



# **STIC Search Report**

## **Biotech-Chem Library**

**STIC Database Tracking Number: 114976**

**TO: Dave Nguyen**  
**Location: REM-2D31**  
**Art Unit: 1632**  
**Thursday, March 04, 2004**

**Case Serial Number: 08/799910**

**From: Edward Hart**  
**Location: Biotech-Chem Library**  
**REM-1A55**  
**Phone: 571-272-2512**

**edward.hart@uspto.gov**

### **Search Notes**

Examiner Nguyen,

Here are the results of the search you requested.

Please feel free to contact me if you have any questions.

Edward Hart

STIC-Biotech/ChemLib

114976

From: Nguyen, Dave  
Sent: Monday, February 23, 2004 6:20 PM  
To: STIC-Biotech/ChemLib  
Subject: sequence search: 08/799,910

Please search amino acids 71-157 of SEQ ID NO: 10, and nucleotide residues 211-468 of SEQ ID NO: 9. Please search all databases including the US pending databases.

Dave Trong Nguyen  
Primary Examiner  
Remsen 2D31  
571-272-0731

Searcher: \_\_\_\_\_  
Phone: \_\_\_\_\_  
Location: \_\_\_\_\_  
Date Picked Up: 2/24/04  
Date Completed: 2/24/04  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH: 1  
NA Sequences: \_\_\_\_\_  
AA Sequences: 1  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST (where applic.)  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys: 051/123  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:20:35 ; Search time 57 Seconds  
(without alignments)  
728.675 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_10\_156  
Perfect score: 764  
Sequence: 1 TMTILQAPTPAPSTIPGRR.....RPSYALDLSTLQKPPAAF 147

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

- 1: Geneseq1980s:\*
- 2: Geneseq1990s:\*
- 3: Geneseq2000s:\*
- 4: Geneseq2001s:\*
- 5: Geneseq2002s:\*
- 6: Geneseq2003as:\*
- 7: Geneseq2003bs:\*
- 8: Geneseq2004s:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description        |
|------------|-------|-------------|--------|-------|--------------------|
| 1          | 764   | 100.0       | 156    | 3     | AAY45017           |
| 2          | 764   | 100.0       | 156    | 5     | AAY45017 Protein e |
| 3          | 764   | 100.0       | 156    | 7     | AAY78232 Human pro |
| 4          | 764   | 100.0       | 156    | 7     | Add46473 Human Pro |
| 5          | 764   | 100.0       | 156    | 7     | Add44772 Human Pro |
| 6          | 764   | 100.0       | 156    | 7     | Add44775 Human Pro |
| 7          | 764   | 100.0       | 156    | 7     | Add44778 Human Pro |
| 8          | 764   | 100.0       | 156    | 7     | Add44778 Human Pro |
| 9          | 757   | 99.1        | 156    | 2     | AAG74873 Human col |
| 10         | 425   | 55.6        | 83     | 4     | AAG74873 Human col |
| 11         | 423   | 55.4        | 82     | 4     | AAG74873 Human col |
| 12         | 281   | 36.8        | 72     | 4     | AAG74873 Human col |
| 13         | 259   | 33.9        | 83     | 4     | AAG74873 Human col |
| 14         | 120.5 | 15.8        | 744    | 6     | AAG74873 Human col |
| 15         | 112.5 | 14.7        | 649    | 4     | ABG7037 Protein e  |
| 16         | 109   | 14.3        | 8991   | 6     | ABG7037 Protein e  |
| 17         | 107.5 | 14.1        | 245    | 4     | ABG7037 Protein e  |
| 18         | 107   | 14.0        | 130    | 3     | ABG7037 Protein e  |
| 19         | 107   | 14.0        | 130    | 3     | ABG7037 Protein e  |
| 20         | 107   | 14.0        | 130    | 3     | ABG7037 Protein e  |
| 21         | 106.5 | 13.9        | 283    | 4     | ABG7037 Protein e  |
| 22         | 106.5 | 13.9        | 283    | 4     | ABG7037 Protein e  |
| 23         | 106.5 | 13.9        | 283    | 4     | ABG7037 Protein e  |
| 24         | 106.5 | 13.9        | 283    | 4     | ABG7037 Protein e  |
| 25         | 106.5 | 13.9        | 283    | 4     | ABG7037 Protein e  |

|    |       |      |      |   |          |
|----|-------|------|------|---|----------|
| 26 | 106.5 | 13.9 | 2972 | 4 | AAB50363 |
| 27 | 106.5 | 13.9 | 3118 | 4 | AAB50362 |
| 28 | 106   | 13.9 | 130  | 3 | AAG47101 |
| 29 | 106   | 13.9 | 131  | 3 | AAG47100 |
| 30 | 105.5 | 13.8 | 441  | 4 | ABB11413 |
| 31 | 104   | 13.6 | 129  | 2 | ABW14573 |
| 32 | 103.5 | 13.5 | 406  | 4 | ABG27250 |
| 33 | 103   | 13.5 | 256  | 4 | ABG27250 |
| 34 | 103   | 13.5 | 538  | 4 | ABG14000 |
| 35 | 103   | 13.5 | 1013 | 4 | ABG14000 |
| 36 | 102   | 13.4 | 708  | 5 | ABB91504 |
| 37 | 102   | 13.4 | 1427 | 6 | ABG21601 |
| 38 | 101.5 | 13.3 | 180  | 2 | AAW14569 |
| 39 | 101   | 13.2 | 180  | 2 | AAW14562 |
| 40 | 101   | 13.2 | 183  | 2 | AAW14570 |
| 41 | 101   | 13.2 | 242  | 6 | ADB08872 |
| 42 | 101   | 13.2 | 355  | 6 | ADB08874 |
| 43 | 100   | 13.1 | 415  | 4 | ASG30150 |
| 44 | 98.5  | 12.9 | 188  | 3 | AAW68732 |
| 45 | 98.5  | 12.9 | 1000 | 6 | ASG25647 |

## ALIGNMENTS

RESULT 1  
AAY45017  
ID AAY45017 standard; protein; 156 AA.

AC AAY45017;

DT 31-MAY-2000 (first entry)

DE Protein encoded by fchd605 gene.

XX fchd605 gene; human; cardiovascular disease; oncogenic disorder;  
KW diabetic retinopathy; fibroproliferative disorder; arteriosclerosis;  
KW TGF-beta signalling pathway; TGF; transforming growth factor;  
KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;  
KW vascularisation; cytostatic; antidiabetic; ophthalmological.

OS Homo sapiens.

PN W0200006206-A1.

PD 10-FEB-2000.

PF 30-JUL-1999; 99WO-US017394.

PR 30-JUL-1998; 98US-00126640.

PA (MILL-) MILLENNIUM PHARM INC.

PI Falb DA;

DR WPI: 2000-205414/18.

XX N-PSDB; AAZ50711.

XX Identifying substances for ameliorating symptoms of fibroproliferative

XX diseases or oncogenic related disorders.

XX Example; Fig 5; 214pp; English.

XX The patent discloses methods for the treatment and diagnosis of  
XX cardiovascular diseases by novel human genes which are differentially  
XX expressed in different cardiovascular disease states. Compositions which  
XX can modify TGF-beta signalling pathway are identified by screening. These  
XX are used therapeutically to treat fibroproliferative and oncogenic  
XX disorders, especially TGF (transforming growth factor)-beta related  
XX disorders, including diabetic retinopathy, arteriosclerosis, pancreatic  
XX cancer, angiogenesis, inflammation, fibrosis, tumour growth and  
XX vascularisation. The present sequence is the protein product of fchd605  
XX gene which is up-regulated in monocytes treated with oxidised LDL (low

CC density lipoprotein) can be used to design cardiovascular disease  
CC treatment strategies. Depending on whether the up-regulation has a  
CC pathogenic or protective effect treatment methods can be designed to  
CC increase or decrease the activity of the protein product of the gene  
XX  
SQ Sequence 156 AA;  
Query Match 100.0%; Score 764; DB 3; Length 156;  
Best Local Similarity 100.0%; Pred. No. 7.1e-64;  
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TMTILOAPTPAPSTIPGRSGPEITFDPLPFAAAGPAGPSASGRHRSRVLYPR 60  
DB 10 TMTILOAPTPAPSTIPGRSGPEITFDPLPFAAAGPAGPSASGRHRSRVLYPR 69  
QY 61 VVRQLPVEEPNPAKRLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120  
DB 70 VVRQLPVEEPNPAKRLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 129  
QY 121 EPNLTSEPSDYALDLSTFLQHPAAF 147  
DB 130 EPNLTSEPSDYALDLSTFLQHPAAF 156  
RESULT 2  
AAU78232  
ID AAU78232 standard; protein; 156 AA.  
XX  
AC AAU78232;  
DT 05-JUN-2002 (first entry)  
XX  
DE Human proliferation regulated gene-1 protein, prg-1/1EX-1.  
XX  
KW Human; MIVR-1; Mechanically Induced Vascular Receptor 1; cytostatic;  
KW cardiant; cerebroprotective; antiarteriosclerotic; cardiac cell;  
KW anti-apoptotic; vascular endothelial cell; cardiac hypertrophy;  
KW myocardial infarction; stroke; arteriosclerosis; heart failure;  
KW proliferation regulated gene-1; prg-1; IEX-1.  
XX  
OS Homo sapiens.  
XX  
PN WO200216415-A2.  
XX  
PD 28-FEB-2002.  
XX  
PF 21-AUG-2001; 2001WO-US026089.  
XX  
PR 22-AUG-2000; 2000US-0227159P.  
XX  
PA (BGMH ) BRIGHAM & WOMENS HOSPITAL INC.  
PA (PFIZ ) PFIZER INC.  
XX  
PI Lee RT, Landschulz KT, Kennedy SP, Thompson JF, Turi TG;  
XX  
DR WPI; 2002-280912/32.  
DR N-PSDB; ABK12138.  
XX  
PT Novel nucleic acid molecule encoding Mechanically Induced Vascular  
PT Receptor-i polypeptide, useful for treating cardiovascular diseases.  
XX  
XX Example; Page 90-91; 105pp; English.  
XX  
CC The invention relates to an isolated nucleic acid molecule encoding a  
CC Mechanically Induced Vascular Receptor (MIVR)-1 polypeptide having  
CC cardiac cell anti-apoptotic activity and fragments of it provided they  
CC are not identical to Genbank sequences A176144.1, A1594390, NM 004338  
CC and A0177461. Also included are expression vectors, host cells, the MIVR-  
CC 1 polypeptide, MIVR-1 binding peptides, modulators of MIVR-1, contacting  
CC a molecule having cardiac cell anti-apoptotic activity with a candidate  
CC agent, where the molecule is a nucleic acid molecule comprising MIVR-1,  
CC IEX-1, VDRP-1, BTG-2 and TIS-1ld or its expression product, determining  
CC if the anti-apoptotic activity is modulated and thereby identifying a

CC modulator. The cardiac cell anti-apoptotic molecules and nucleic acids  
CC of the invention are useful for treating, diagnosing and monitoring  
CC progression of such diseases and disorders as characterised by increased  
CC apoptotic cell-death of vascular endothelial cells e.g. cardiac  
CC hypertrophy, myocardial infarction, stroke, arteriosclerosis and heart  
CC failure. The present sequence represents human proliferation regulated  
CC gene-1, prg-1 (also known as IEX-1) protein which is also mechanically  
CC induced and has an apoptosis regulatory function  
XX  
SQ Sequence 156 AA;  
Query Match 100.0%; Score 764; DB 5; Length 156;  
Best Local Similarity 100.0%; Pred. No. 7.1e-64;  
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TMTILOAPTPAPSTIPGRSGPEITFDPLPFAAAGPAGPSASGRHRSRVLYPR 60  
DB 10 TMTILOAPTPAPSTIPGRSGPEITFDPLPFAAAGPAGPSASGRHRSRVLYPR 69  
QY 61 VVRQLPVEEPNPAKRLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120  
DB 70 VVRQLPVEEPNPAKRLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 129  
QY 121 EPNLTSEPSDYALDLSTFLQHPAAF 147  
DB 130 EPNLTSEPSDYALDLSTFLQHPAAF 156  
RESULT 3  
ADD46473  
ID ADD46473 standard; protein; 156 AA.  
XX  
AC ADD46473;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Human Protein CAA65304, SEQ ID NO 12154.  
XX  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
PF 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEMO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
XX WPI; 2003-268312/26.  
DR GENBANK; CAA65304.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a



CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 156 AA;

Query Match 100.0%; Score 764; DB 7; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-64;  
 Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTPAPSTIPGRSGPEIFTFDLPPEPAAAPAGPSASRGHRKRSRVLYPR 60  
 DB 10 TMTILQAPTPAPSTIPGRSGPEIFTFDLPPEPAAAPAGPSASRGHRKRSRVLYPR 69  
 QY 61 VVRQLPVEEPNPAKRLFLLLTIVFCQLMAEGVPAIPPEPAPNAASLAPTPVSPVL 120  
 DB 70 VVRQLPVEEPNPAKRLFLLLTIVFCQLMAEGVPAIPPEPAPNAASLAPTPVSPVL 129  
 QY 121 EPNLTSEPSDYALDLSTFLQQHPAAF 147  
 DB 130 EPNLTSEPSDYALDLSTFLQQHPAAF 156

RESULT 4  
 ID ADD44772 standard; protein; 156 AA.  
 AC ADD44772;  
 DT 29-JAN-2004 (first entry)  
 DE Human Protein CAA65304, SEQ ID NO 10203.  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX Homo sapiens.  
 XX WO2003016475-A2.  
 XX 27-FEB-2003.  
 XX 14-AUG-2002; 2002WO-US025765.  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX (GEHO) GEN HOSPITAL CORP.  
 PA (FARR) BAYER AG.  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-266312/26.

DR GENBANK; CAA65304.  
 XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX Claim 1; Page: 10:7pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (SNI), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 XX Sequence 156 AA;

Query Match 100.0%; Score 764; DB 7; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-64;  
 Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTPAPSTIPGRSGPEIFTFDLPPEPAAAPAGPSASRGHRKRSRVLYPR 60  
 DB 10 TMTILQAPTPAPSTIPGRSGPEIFTFDLPPEPAAAPAGPSASRGHRKRSRVLYPR 69  
 QY 61 VVRQLPVEEPNPAKRLFLLLTIVFCQLMAEGVPAIPPEPAPNAASLAPTPVSPVL 120  
 DB 70 VVRQLPVEEPNPAKRLFLLLTIVFCQLMAEGVPAIPPEPAPNAASLAPTPVSPVL 129  
 QY 121 EPNLTSEPSDYALDLSTFLQQHPAAF 147  
 DB 130 EPNLTSEPSDYALDLSTFLQQHPAAF 156

RESULT 5  
 ID ADD44775 standard; protein; 156 AA.  
 AC ADD44775;  
 DT 29-JAN-2004 (first entry)  
 DE Human Protein CAA65304, SEQ ID NO 10206.  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX Homo sapiens.  
 XX WO2003016475-A2.  
 XX 27-FEB-2003.

XX 14-AUG-2002; 2002WO-US025765.  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX (GHEO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 DR GENBANK; CAA65304.  
 XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX Claim 1; Page; 1017pp; English.  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 156 AA;  
 CC  
 CC Query Match 100.0%; Score 764; DB 7; Length 156;  
 CC Best Local Similarity 100.0%; Pred. No. 7.le-64;  
 CC Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWTLQAPTAPSTIPGRSGPEFTFDLPPEAAAPAGPSASGRKRRVLYPR 60  
 DB 10 TWTLQAPTAPSTIPGRSGPEFTFDLPPEAAAPAGPSASGRKRRVLYPR 69  
 QY 61 VVRQLFVEENPAKRLIFLLTIVFCQILMAEGVPAIPPEPADNAPASLAPTPSPVL 120  
 DB 70 VVRQLFVEENPAKRLIFLLTIVFCQILMAEGVPAIPPEPADNAPASLAPTPSPVL 129  
 QY 121 EPNLSEPSDYDLSTFLQHPAF 147  
 DB 130 EPNLSEPSDYDLSTFLQHPAF 156  
 RESULT 6  
 ADD44769  
 ID ADD44769 standard; protein; 156 AA.  
 XX  
 AC ADD44769;  
 XX

DT 29-JAN-2004 (first entry)  
 XX Human Protein CAA65304, SEQ ID NO 10200.  
 XX Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX Homo sapiens.  
 XX WO2003016475-A2.  
 XX 27-FEB-2003.  
 XX 14-AUG-2002; 2002WO-US025765.  
 XX 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX (GHEO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 XX WPI; 2003-268312/26.  
 DR GENBANK; CAA65304.  
 XX New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX Claim 1; Page; 1017pp; English.  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 156 AA;  
 CC  
 CC Query Match 100.0%; Score 764; DB 7; Length 156;  
 CC Best Local Similarity 100.0%; Pred. No. 7.le-64;  
 CC Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWTLQAPTAPSTIPGRSGPEFTFDLPPEAAAPAGPSASGRKRRVLYPR 60  
 DB 10 TWTLQAPTAPSTIPGRSGPEFTFDLPPEAAAPAGPSASGRKRRVLYPR 69  
 QY 61 VVRQLFVEENPAKRLIFLLTIVFCQILMAEGVPAIPPEPADNAPASLAPTPSPVL 120  
 DB 70 VVRQLFVEENPAKRLIFLLTIVFCQILMAEGVPAIPPEPADNAPASLAPTPSPVL 129  
 QY 121 EPNLSEPSDYDLSTFLQHPAF 147  
 DB 130 EPNLSEPSDYDLSTFLQHPAF 156  
 RESULT 6  
 ADD44769  
 ID ADD44769 standard; protein; 156 AA.  
 XX  
 AC ADD44769;  
 XX

Db 70 VVRQLPVEENPAKRLIFLLITIVFCQILMAEGVPAFLPEDAPNAASIAFTPVSPVL 129

QY 121 EPNLTSEPSDYALDLSSTFLQHPAAF 147

Db 130 EPNLTSEPSDYALDLSSTFLQHPAAF 156

RESULT 7

ADD44778

ID ADD44778 standard; protein; 156 AA.

XX

AC ADD44778;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human Protein CAA65304, SEQ ID NO 10209.

XX

KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

XX

OS Homo sapiens.

XX

PN WO2003016475-A2.

PD 27-FEB-2003.

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

PA (GEMO ) GEN HOSPITAL CORP.

PA (FARB ) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

DR GENBANK; CAA65304.

XX

PT New composition comprising two or more isolated polypeptides, useful for

PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat

CC or human polynucleotides or a polynucleotide which represents a fragment,

CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell

CC comprising the vector, a method for identifying a nucleotide sequence

CC which is differentially regulated in an animal subjected to pain and a

CC kit to perform the method, an array, a method for identifying an agent

CC that increases or decreases the expression of the polynucleotide sequence

CC that is differentially expressed in neuronal tissue of a first animal

CC subjected to pain, a method for identifying a compound which regulates

CC the expression of a polynucleotide sequence which is differentially

CC expressed in an animal subjected to pain, a method for identifying a

CC compound that regulates the activity of one or more of the

CC polynucleotides, a method for producing a pharmaceutical composition, a

CC method for identifying a compound or small molecule that regulates the

CC activity in an animal of one or more of the polypeptides given in the

CC specification, a method for identifying a compound useful in treating

CC pain and a pharmaceutical composition comprising the one or more

CC polypeptides or their antibodies. The polynucleotide or the compound that

CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction

CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene

CC therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic form directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 156 AA;

Query Match 100.0%; Score 764; DB 7; Length 156;

Best Local Similarity 100.0%; Pred. No. 7.le-64;

Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTPAPSTIPGRRGSGFEIFTFDLPPEPAAAPAGRPSASGRKRSRVLYPR 60

Db 10 TMTILQAPTPAPSTIPGRRGSGFEIFTFDLPPEPAAAPAGRPSASGRKRSRVLYPR 69

QY 61 VVRQLPVEENPAKRLIFLLITIVFCQILMAEGVPAFLPEDAPNAASIAFTPVSPVL 120

Db 70 VVRQLPVEENPAKRLIFLLITIVFCQILMAEGVPAFLPEDAPNAASIAFTPVSPVL 129

QY 121 EPNLTSEPSDYALDLSSTFLQHPAAF 147

Db 130 EPNLTSEPSDYALDLSSTFLQHPAAF 156

RESULT 8

AAG74873

ID AAG74873 standard; protein; 166 AA.

XX

AC AAG74873;

XX

DT 03-SEP-2001 (first entry)

XX

XX Human colon cancer antigen protein SEQ ID NO:5637.

XX

DE Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 6.

XX

OS Homo sapiens.

XX

PN WO200122920-A2.

XX

PD 05-APR-2001.

XX

PF 28-SEP-2000; 2000WO-US026524.

XX

PR 29-SEP-1999; 99US-0157137P.

PR 03-NOV-1999; 99US-0163280P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Ruben SM, Barash SC, Birse CE, Rosen CA;

XX

DR WPI; 2001-235357/24.

DR N-PSDB; AAH34278.

XX

PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers.

XX

PS Claim 11; Page 7188-7189; 9803pp; English.

XX

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where the

CC proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene therapy

CC and vaccine production. N and P may be used in the prevention, diagnosis

CC and treatment of diseases associated with inappropriate P expression. For

CC example, N and P may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of P by expressing inactive proteins or to

CC supplement the patients own production of P. Additionally, N may be used

CC to produce the colon cancer-associated ps, by inserting the nucleic acids

CC into a host cell and culturing the cell to express the proteins. N and P

CC can be used in the prevention, diagnosis and treatment of colorectal

CC carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent

CC sequences used in the exemplification of the present invention. N.B.

CC Pages 666 to 682 and page 7053 of the sequence listing were missing at

CC time of publication, meaning no sequences are present for SEQ ID NO:1027

CC to 1052, 7921 and 7922  
 XX Sequence 156 AA;  
 SQ Query Match 100.0%; Score 764; DB 4; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-64;  
 Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TWTLQAPTAPSTIFGPRGSGPEIFTFDPLPEPAAAPAGPSASGRGKSRRLVLYPR 60  
 Db 20 TWTLQAPTAPSTIFGPRGSGPEIFTFDPLPEPAAAPAGPSASGRGKSRRLVLYPR 79  
 QY 61 VVRQLPVEENPAKLLLELLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120  
 Db 80 VVRQLPVEENPAKLLLELLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 139  
 QY 121 EPNLTSEPSDYALDLSTFLOQHPAAF 147  
 Db 140 EPNLTSEPSDYALDLSTFLOQHPAAF 166  
 RESULT 9  
 AAW36006  
 ID AAW36006 standard; protein; 156 AA.  
 XX  
 AC AAW36006;  
 XX  
 DT 03-MAR-1998 (first entry)  
 XX  
 DE Human Fchd605 gene product.  
 XX  
 KW Fchd605 gene; differential expression; monocyte; human; foam cell;  
 KW cardiovascular disease; atherosclerosis; ischaemia; reperfusion;  
 KW hypertension; restenosis; arterial inflammation; therapy; diagnosis;  
 KW drug screening; marker.  
 OS  
 CS Homo sapiens.  
 XX  
 PN WO9730065-AL.  
 PD 21-AUG-1997.  
 PF 14-FEB-1997; 97WO-US002291.  
 PR 16-FEB-1996; 96US-0011787P.  
 PR 13-FEB-1997; 97US-00799910.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 XX  
 PI Falb DA;  
 XX  
 DR WPI; 1997-424966/39.  
 DR N-PSDE; AAT94471.  
 XX  
 PT New genes differentially expressed in cardiovascular disease - used for  
 PT diagnosis, drug screening and treatment of cardiovascular disease, e.g.  
 PT atherosclerosis, restenosis, hypertension, etc.  
 XX  
 PS Example 6; Fig 5; 163pp; English.  
 XX  
 CC This protein is encoded by the novel human fchd605 gene (see AAT94471)  
 CC that is up-regulated in monocytes treated with oxidised low density  
 CC lipoproteins that simulate the conditions under which foam cells develop  
 CC during atherogenesis. The protein has sequence similarity to the mouse  
 CC gily96 gene and to ESR AAT94532. Novel fchd531, fchd540, fchd545, fchd502  
 CC and fchd605 genes (see AAT94467-71) provide a fingerprint for the study  
 CC of cardiovascular diseases, including atherosclerosis.  
 CC ischaemia/reperfusion, hypertension, restenosis and arterial  
 CC inflammation. Methods are provided for the diagnosis and monitoring in  
 CC clinical trials, screening for therapeutically effective compounds, and  
 CC treatment of cardiovascular diseases based on discoveries regarding the  
 CC expression patterns of these novel genes  
 XX

SQ Sequence 156 AA;  
 Query Match 99.1%; Score 757; DB 2; Length 156;  
 Best Local Similarity 99.3%; Pred. No. 3.2e-63;  
 Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 TWTLQAPTAPSTIFGPRGSGPEIFTFDPLPEPAAAPAGPSASGRGKSRRLVLYPR 60  
 Db 10 TWTLQAPTAPSTIFGPRGSGPEIFTFDPLPEPAAAPAGPSASGRGKSRRLVLYPR 69  
 QY 61 VVRQLPVEENPAKLLLELLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120  
 Db 70 VVRQLPVEENPAKLLLELLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 129  
 QY 121 EPNLTSEPSDYALDLSTFLOQHPAAF 147  
 Db 130 EPNLTSEPSDYALDLSTFLOQHPAAF 156  
 RESULT 10  
 AAE03976  
 ID AAE03976 standard; peptide; 83 AA.  
 XX  
 AC AAE03976;  
 XX  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE Human gene 36 encoded secreted protein fragment, SEQ ID NO:168.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification.  
 OS  
 CS Homo sapiens.  
 XX  
 PN WO200077022-AL.  
 XX  
 PD 21-DEC-2000.  
 XX  
 PF 01-JUN-2000; 2000WO-US015136.  
 PR 11-JUN-1999; 99US-0138629P.  
 XX  
 PA (HUNA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis GA;  
 XX  
 DR WPI; 2001-367020/38.  
 XX  
 PT Nucleic acids encoding 50 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating diseases, e.g. Parkinson's  
 PT disease, botulism, cancers and Scimitar syndrome.  
 XX  
 PS Disclosure; Page 591; 614pp; English.  
 XX  
 CC AAD08345-AAD08394 represent cDNAs corresponding to 50 human secreted  
 CC protein genes and AAE03898-AAE03947 represent the proteins they encode.  
 CC AAE03948-AAE03996 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 50 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC

CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein fragment referred to  
 CC in the disclosure of the invention  
 XX  
 SQ Sequence 83 AA;

Query Match 55.6%; Score 425; DB 4; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;  
 Matches 83; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 61 VVRQLVPEPNPAKLLFLLLTTFVFCQILMAEGVPAPLPEDAPNAASLAFTVPSPVL 120  
 Db 1 VVRQLVPEPNPAKLLFLLLTTFVFCQILMAEGVPAPLPEDAPNAASLAFTVPSPVL 60  
 QY 121 EPPNLTSEPSDYALDLSFLQOH 143  
 Db 61 EPPNLTSEPSDYALDLSFLQOH 83

RESULT 11  
 AAE03933  
 ID AAE03933 standard; protein; 82 AA.  
 AC AAE03933;  
 XX  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE Human gene 36 encoded secreted protein HAUC84, SEQ ID NO:96.

XX Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnerability; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification.  
 XX  
 OS Homo sapiens.

XX Key Location/Qualifiers  
 FH Peptide 1..33  
 FT /label= signal\_peptide  
 FT Protein 34..82  
 FT /note= "Mature secreted protein"  
 XX  
 PN WD200077022-A1.  
 XX  
 PD 21-DEC-2000.  
 XX  
 PF 01-JUN-2000; 2000WO-US015136.  
 XX  
 PR 11-JUN-1999; 99US-0138629P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX

PI Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI; 2001-367020/38.  
 DR N-PSDB; AAD08380.  
 XX  
 PT Nucleic acids encoding 50 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating diseases, e.g. Parkinson's  
 PT disease, botulism, cancers and Scimitar syndrome.  
 XX  
 PS Claim 11; Page 546; 614pp; English.

XX AAD08345-AAD08394 represent cDNAs corresponding to 50 human secreted  
 CC protein genes and AAE03898-AAE03947 represent the proteins they encode.  
 CC AAE03948-AAE03996 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 50 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies, auto-  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein of the invention  
 XX  
 SQ Sequence 82 AA;

Query Match 55.4%; Score 423; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 3.5e-32;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 66 LPVERPNPAKLLFLLLTTFVFCQILMAEGVPAPLPEDAPNAASLAFTVPSPVLEPNL 125  
 Db 1 LPVERPNPAKLLFLLLTTFVFCQILMAEGVPAPLPEDAPNAASLAFTVPSPVLEPNL 60  
 QY 126 TSEPSDYALDLSFLQOHPAAP 147  
 Db 61 TSEPSDYALDLSFLQOHPAAP 82

RESULT 12  
 AAE03977  
 ID AAE03977 standard; peptide; 72 AA.  
 XX  
 AC AAE03977;  
 XX  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE Human gene 36 encoded secreted protein fragment, SEQ ID NO:170.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;

PN WO200277183-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US009107.  
 XX 21-MAR-2001; 2001US-00815242.  
 PR 06-SEP-2001; 2001US-00948993.  
 PR 25-OCT-2001; 2001US-0342923P.  
 PR 08-FEB-2002; 2002US-00072851.  
 PR 06-MAR-2002; 2002US-0362699P.  
 XX (ELIT-) ELITRA PHARM INC.  
 XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
 XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
 DR WPI; 2003-029926/02.  
 DR N-PSDB; ACA27157.  
 XX New antisense nucleic acids, useful for identifying proteins or screening  
 PT for homologous nucleic acids required for cellular proliferation to  
 PT isolate candidate molecules for rational drug discovery programs.  
 XX Claim 25; SEQ ID NO 51211; 1766pp; English.  
 XX The invention relates to an isolated nucleic acid comprising any one of  
 CC the 6213 antisense sequences given in the specification where expression  
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
 CC (1) a vector comprising a promoter operably linked to the nucleic acid  
 CC encoding a polypeptide whose expression is inhibited by the antisense  
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
 CC polypeptide or its fragment whose expression is inhibited by the  
 CC antisense nucleic acid; (4) an antibody capable of specifically binding  
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
 CC proliferation or the activity of a gene in an operon required for  
 CC the gene product or that has an activity against a biological pathway;  
 CC required for proliferation, or that inhibits cellular proliferation; (8)  
 CC identifying a gene required for cellular proliferation or the biological  
 CC pathway in which a proliferation-required gene or its gene product lies  
 CC or a gene on which the test compound that inhibits proliferation of an  
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
 CC compound's activity; (11) a culture comprising strains in which the gene  
 CC product is overexpressed or underexpressed; (12) determining the extent  
 CC to which each of the strains is present in a culture or collection of  
 CC strains; or (13) identifying the target of a compound that inhibits the  
 CC proliferation of an organism. The antisense nucleic acids are useful for  
 CC identifying proteins or screening for homologous nucleic acids required  
 CC for cellular proliferation to isolate candidate molecules for rational  
 CC drug discovery programs, or for screening homologous nucleic acids  
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
 CC the target prokaryotic essential genes. Note: The sequence data for this  
 CC patent did not form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 744 AA;  
 SQ  
 Query Match 15.8%; Score 120.5; DB 6; Length 744;  
 Best Local Similarity 30.8%; Pred. No. 0.011;  
 Matches 40; Conservative 5; Mismatches 38; Indels 47; Gaps 5;  
 QY 8 PTPASTIP-----GPRGSGEFTFDPPEFAAPAGRSAGRHKRS 53  
 DB 442 PTPAALAPQAAAPRAVAKSKFPQAPPE-----PEPEAPEAPATVAAT----- 490  
 QY 54 RRVLYPRVVRQLFVEBPNPAAKLLFLLLTVFQILMAEGVPAAPPP-EDAPNAASLA 112  
 DB 491 -----PPAVAFPEFVAPAAQAR-----AAEFAPAPVPPNEDLPEAVA-A 529  
 QY 113 PTPVSVLEP 122

Db 530 PAPAQFAAAP 539  
 RESULT 15  
 ABB67037  
 ID ABB67037 standard; protein, 649 AA.  
 AC ABB67037;  
 DT 26-MAR-2002 (first entry)  
 XX Drosophila melanogaster polypeptide SEQ ID NO 27903.  
 XX Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX 23-MAR-2001; 2001WO-US009231.  
 PR 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX (PEKE ) PE CORP NY.  
 XX Venter JC, Adams M, Li PWD, Myers EW;  
 WPI; 2001-656860/75.  
 DR N-PSDB; ABL11140.  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX Disclosure; SEQ ID NO 27903; 21pp + Sequence Listing; English.  
 XX The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABBS7737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX Sequence 649 AA;  
 SQ  
 Query Match 14.7%; Score 112.5; DB 4; Length 649;  
 Best Local Similarity 25.8%; Pred. No. 0.053;  
 Matches 46; Conservative 26; Mismatches 67; Indels 39; Gaps 7;  
 QY 1 TWILQA--PTAPSTIPGRRSGEFTFDPPEFAAPAGRSAGRHKRSRVLY 58  
 DB 342 TIEVLOSIFETPAP---PPPKTAPPELPVLKPTTEEDPTPOLRLTSDNEHRKSRVTR 398  
 QY 59 P-----RVVRQLFVEBPNPAAKLLFLLL-----TIVFQILMAEGVPAAP 99  
 DB 399 PLQSLETFKMLLVNAQLYKENVPHPKLLIGLKXDEDEEHT-----VLIPQPKRSP 453  
 QY 100 LPPEAPNAASLAIPVSPVLEPNITSEPSDY-----ALDSTFLQHQHAP 147  
 DB 454 LTPKDVVDYFPRTKIVNGIQDDFAGRTVFFGWIATITMRMLSLSTPCVFPKAP 511  
 Search completed: February 25, 2004, 05:26:43  
 Job time : 61 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:25:35 ; Search time 23 Seconds  
(without alignments)  
329.957 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_10\_156

Perfect score: 764  
Sequence: 1 TWILQAPTPASIPGRR.....PESYALDLSTFIQHPAP 147

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA\*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pcp:\*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pcp:\*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pcp:\*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pcp:\*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pcp:\*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pcp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description          |
|------------|-------|-------------|--------|-------|----------------------|
| 1          | 764   | 100.0       | 156    | 3     | US-08-826-246-10     |
| 2          | 764   | 100.0       | 156    | 3     | US-08-944-495-10     |
| 3          | 764   | 100.0       | 156    | 3     | US-09-126-640-11     |
| 4          | 764   | 100.0       | 156    | 3     | US-08-925-588-10     |
| 5          | 764   | 100.0       | 156    | 4     | US-09-288-292A-11    |
| 6          | 764   | 100.0       | 156    | 4     | US-09-372-044-10     |
| 7          | 764   | 100.0       | 156    | 4     | US-08-825-486-10     |
| 8          | 109   | 14.3        | 8991   | 4     | US-08-714-741-32     |
| 9          | 107   | 14.0        | 206    | 4     | US-08-529-055-54     |
| 10         | 106.5 | 13.9        | 2972   | 4     | US-09-579-181-2      |
| 11         | 106.5 | 13.9        | 3118   | 4     | US-09-579-181-1      |
| 12         | 104   | 13.6        | 129    | 4     | US-08-529-055-53     |
| 13         | 102.5 | 13.4        | 361    | 4     | US-09-510-031A-5     |
| 14         | 101.5 | 13.3        | 193    | 4     | US-08-529-055-49     |
| 15         | 101   | 13.2        | 181    | 4     | US-08-529-055-42     |
| 16         | 101   | 13.2        | 183    | 4     | US-08-529-055-50     |
| 17         | 98.5  | 12.9        | 655    | 4     | US-09-252-991A-31645 |
| 18         | 97    | 12.7        | 493    | 4     | US-09-252-991A-31227 |
| 19         | 96    | 12.6        | 805    | 4     | US-08-714-741-46     |
| 20         | 95.5  | 12.5        | 739    | 3     | US-09-035-648-24     |
| 21         | 95.5  | 12.5        | 739    | 3     | US-09-001-951-24     |
| 22         | 95.5  | 12.5        | 739    | 3     | US-08-818-829-24     |
| 23         | 95    | 12.4        | 429    | 4     | US-09-252-991A-16841 |
| 24         | 94    | 12.3        | 815    | 4     | US-09-266-225D-12    |
| 25         | 93.5  | 12.2        | 185    | 4     | US-08-529-055-46     |
| 26         | 93    | 12.2        | 208    | 4     | US-09-252-991A-17849 |
| 27         | 92.5  | 12.1        | 558    | 4     | US-09-252-991A-17202 |

|    |      |      |      |   |                      |
|----|------|------|------|---|----------------------|
| 28 | 92.5 | 12.1 | 1596 | 3 | US-09-356-952-3      |
| 29 | 92   | 12.0 | 166  | 4 | US-08-539-055-48     |
| 30 | 92   | 12.0 | 166  | 4 | US-08-529-055-55     |
| 31 | 92   | 12.0 | 1400 | 1 | US-08-080-255-7      |
| 32 | 92   | 12.0 | 1400 | 3 | US-08-465-713-7      |
| 33 | 92   | 12.0 | 1400 | 5 | PCT-US93-05857-7     |
| 34 | 92   | 12.0 | 3969 | 3 | US-08-061-376-5      |
| 35 | 91.5 | 12.0 | 615  | 3 | US-08-676-444-44     |
| 36 | 91   | 11.9 | 663  | 4 | US-09-252-991A-30843 |
| 37 | 90.5 | 11.8 | 424  | 4 | US-09-252-991A-30209 |
| 38 | 90.5 | 11.8 | 543  | 4 | US-09-252-991A-17787 |
| 39 | 90   | 11.8 | 152  | 4 | US-09-214-909-22     |
| 40 | 90   | 11.8 | 195  | 4 | US-09-252-991A-20967 |
| 41 | 90   | 11.8 | 254  | 4 | US-09-216-393B-126   |
| 42 | 90   | 11.8 | 279  | 4 | US-09-252-991A-23821 |
| 43 | 90   | 11.8 | 1027 | 4 | US-09-252-991A-17886 |
| 44 | 89.5 | 11.7 | 171  | 4 | US-09-216-393B-71    |
| 45 | 89.5 | 11.7 | 171  | 4 | US-09-216-393B-274   |

## ALIGNMENTS

RESULT 1  
US-08-826-246-10  
; Sequence 10, Application US/08826246  
; Patent No. 6048749  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-826-246-10

Sequence 3, Appli  
Sequence 48, Appl  
Sequence 55, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 7, Appl  
Sequence 5, Appl  
Sequence 44, Appl  
Sequence 30843, A  
Sequence 30209, A  
Sequence 17787, A  
Sequence 22, Appl  
Sequence 20967, A  
Sequence 126, App  
Sequence 23821, A  
Sequence 17886, A  
Sequence 71, Appl  
Sequence 274, App



```
Query Match      100.0%; Score 764; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWTILOAPTAPSTIPGPRGSGPEIFTFDLPPEPAAAPAGPSASRGHRSRRVLYPR 60
DB 10 TWTILOAPTAPSTIPGPRGSGPEIFTFDLPPEPAAAPAGPSASRGHRSRRVLYPR 69

QY 61 VVRQLPVEEPNPAKLLFLLLTIVFCQILMAEBGVPAFLPPEDAPNAASLAPTVPSPVL 120
DB 70 VVRQLPVEEPNPAKLLFLLLTIVFCQILMAEBGVPAFLPPEDAPNAASLAPTVPSPVL 129

QY 121 EPNLTSPSDYALDLSTFLQHPAAF 147
DB 130 EPNLTSPSDYALDLSTFLQHPAAF 156

RESULT 2
US-08-944-495-10
; Sequence 10, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,495
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-944-495-10

Query Match      100.0%; Score 764; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWTILOAPTAPSTIPGPRGSGPEIFTFDLPPEPAAAPAGPSASRGHRSRRVLYPR 60
DB 10 TWTILOAPTAPSTIPGPRGSGPEIFTFDLPPEPAAAPAGPSASRGHRSRRVLYPR 69

QY 61 VVRQLPVEEPNPAKLLFLLLTIVFCQILMAEBGVPAFLPPEDAPNAASLAPTVPSPVL 120
```

```
DB 70 VVRQLPVEEPNPAKLLFLLLTIVFCQILMAEBGVPAFLPPEDAPNAASLAPTVPSPVL 129
QY 121 EPNLTSPSDYALDLSTFLQHPAAF 147
DB 130 EPNLTSPSDYALDLSTFLQHPAAF 156

RESULT 3
US-09-126-640-11
; Sequence 11, Application US/09126640A
; Patent No. 609823
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean A
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/126,640A
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 08/870,434
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 08/799,910
; EARLIER FILING DATE: 1997-02-13
; EARLIER APPLICATION NUMBER: 60/011,787
; EARLIER FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-126-640-11

Query Match      100.0%; Score 764; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWTILOAPTAPSTIPGPRGSGPEIFTFDLPPEPAAAPAGPSASRGHRSRRVLYPR 60
DB 10 TWTILOAPTAPSTIPGPRGSGPEIFTFDLPPEPAAAPAGPSASRGHRSRRVLYPR 69

QY 61 VVRQLPVEEPNPAKLLFLLLTIVFCQILMAEBGVPAFLPPEDAPNAASLAPTVPSPVL 120
DB 70 VVRQLPVEEPNPAKLLFLLLTIVFCQILMAEBGVPAFLPPEDAPNAASLAPTVPSPVL 129

QY 121 EPNLTSPSDYALDLSTFLQHPAAF 147
DB 130 EPNLTSPSDYALDLSTFLQHPAAF 156

RESULT 4
US-08-925-598-10
; Sequence 10, Application US/08925598
; Patent No. 6221628
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
```

```

; APPLICATION NUMBER: US/08/925,588
; FILING DATE: 08-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-925-588-10

Query Match          100.0%; Score 764; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTPAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGPSASRGHKKRSRVLYPR 60
DB 10 TMTILQAPTPAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGPSASRGHKKRSRVLYPR 69

QY 61 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120
DB 70 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 129

QY 121 EPNLTSEPSDYALDLSTFLQQHPAAF 147
DB 130 EPNLTSEPSDYALDLSTFLQQHPAAF 156

RESULT 5
US-09-288-292A-11
; Sequence 11, Application US/09288292A
; Patent No. 6355194
; GENERAL INFORMATION:
; APPLICANT: Dean A. Faib
; APPLICANT: Katherine Galvin
; APPLICANT: Michael Donovan
; APPLICANT: Dennis Huszar
; APPLICANT: Michael A. Gimbrone, Jr.
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Diagnosis of
; FILE REFERENCE: 7853-140-999
; CURRENT APPLICATION NUMBER: US/09/288,292A
; CURRENT FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 08/870,434
; PRIOR FILING DATE: 1997-06-06
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; PRIOR APPLICATION NUMBER: 08/485,573
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/386,844
; PRIOR FILING DATE: 1995-02-10
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Past-Seq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens

```

```

US-09-288-292A-11
Query Match          100.0%; Score 764; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTPAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGPSASRGHKKRSRVLYPR 60
DB 10 TMTILQAPTPAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGPSASRGHKKRSRVLYPR 69

QY 61 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120
DB 70 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 129

QY 121 EPNLTSEPSDYALDLSTFLQQHPAAF 147
DB 130 EPNLTSEPSDYALDLSTFLQQHPAAF 156

RESULT 6
US-09-372-044-10
; Sequence 10, Application US/09372044A
; Patent No. 6492126
; GENERAL INFORMATION:
; APPLICANT: Dean Faib et al.
; TITLE OF INVENTION: Compositions and Methods for the
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Past-Seq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-372-044-10

Query Match          100.0%; Score 764; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.4e-69;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTPAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGPSASRGHKKRSRVLYPR 60
DB 10 TMTILQAPTPAPSTIPGRRGSGPEIFTFDPLPEPAAAPAGPSASRGHKKRSRVLYPR 69

QY 61 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 120
DB 70 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVL 129

QY 121 EPNLTSEPSDYALDLSTFLQQHPAAF 147
DB 130 EPNLTSEPSDYALDLSTFLQQHPAAF 156

RESULT 7
US-08-825-486-10
; Sequence 10, Application US/08825486
; Patent No. 6534641
; GENERAL INFORMATION:
; APPLICANT: Faib, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 7853-140-999
; CURRENT APPLICATION NUMBER: US/08/825,486
; CURRENT FILING DATE: 1999-04-08
; PRIOR APPLICATION NUMBER: 08/799,910
; PRIOR FILING DATE: 1997-02-13
; PRIOR APPLICATION NUMBER: 60/011,787
; PRIOR FILING DATE: 1996-02-16
; PRIOR APPLICATION NUMBER: 08/485,573
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 08/386,844
; PRIOR FILING DATE: 1995-02-10
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Past-Seq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens

```

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/825,486  
FILING DATE: 28-MAR-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/799,910  
FILING DATE: 13-FEB-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A  
REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7853-077-999  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)7909090  
TELEFAX: (212)8699741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 156 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-825-486-10

Query Match 100.0%; Score 764; DB 4; Length 156;  
Best Local Similarity 100.0%; Pred. No. 1.4e-69; Indels 0; Gaps 0;  
Matches 147; Conservative 0; Mismatches 0;  
QY 1 TWTLQAPTAPSTIPGRGSGPEIFTFDLPPEAAAPAGPSASGRHKSRRLVYPR 60  
DB 10 TWTLQAPTAPSTIPGRGSGPEIFTFDLPPEAAAPAGPSASGRHKSRRLVYPR 69  
QY 61 VVRQLPVRENPAKRLFLTLTVFCQILMAEAGVPALPPEDAPNAASLAPTPVSVL 120  
DB 70 VVRQLPVRENPAKRLFLTLTVFCQILMAEAGVPALPPEDAPNAASLAPTPVSVL 129  
QY 121 EPNLTSPSYALDLSTFLQHPA 147  
DB 130 EPNLTSPSYALDLSTFLQHPA 156

RESULT 8  
US-08-714-741-32  
Sequence 32, Application US/08714741  
Patent No. 6500613  
GENERAL INFORMATION:  
APPLICANT: Briles, David E.  
APPLICANT: McDaniel, Larry S.  
APPLICANT: Swiatlo, Edwin  
APPLICANT: Yother, Janet  
APPLICANT: Crain, Marilyn J.  
APPLICANT: Hollingshead, Susan  
APPLICANT: Tart, Rebecca  
APPLICANT: Brooks-Walter, Alexis  
TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,  
TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,  
TITLE OF INVENTION: PORTIONS AND PRODUCTS  
NUMBER OF SEQUENCES: 47  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/714,741  
FILING DATE: 16-SEP-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer Esq., William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454312-2460  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 32:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8991 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: amino acid  
US-08-714-741-32

Query Match 14.3%; Score 109; DB 4; Length 8991;  
Best Local Similarity 27.1%; Pred. No. 0.13;  
Matches 35; Conservative 11; Mismatches 31; Indels 52; Gaps 6;  
QY 7 APTAPSTIPGRGSGPEIFTFDLPPEAAAPAGPSASGRHKSRRLVYPRVVRQL 66  
DB 5822 APAPAPAPAPKAPAPK-----PAPAPAPAPAPK----- 5854  
QY 67 PVEENPAKRLFLTLTVFCQILMAEAGVPALP-PEDAPNAASLAPTPVSP-VLEPFN 124  
DB 5855 PAKPAPA-----PAPAPKPEKPAKPAKPEKPTKTLKDD 5892  
QY 125 LTSEPSDYA 133  
DB 5893 -ESDSEDYA 5900

RESULT 9  
US-08-529-055-54  
Sequence 54, Application US/08529055  
Patent No. 6592876  
GENERAL INFORMATION:  
APPLICANT: Briles, David E.  
APPLICANT: McDaniel, Larry S.  
APPLICANT: Swiatlo, Edwin  
APPLICANT: Yother, Janet  
APPLICANT: Brooks-Walter, Alexis  
TITLE OF INVENTION: Pneumococcal Genes, Portions  
TITLE OF INVENTION: Thereof, Expression Products  
TITLE OF INVENTION: Therefrom, and Uses of Such Genes,  
TITLE OF INVENTION: Portions and Products  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/529,055  
FILING DATE: 15-SEP-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506







GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:28:10 : Search time 34 Seconds

(without alignments)

912.927 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_10\_156

Perfect score: 764

Sequence: 1 RTWILQAPTAPETIPGRR.....EPDVALDLSTFLQHPAAF 147

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubaa/PCT NEW PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubaa/US06\_NEW PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubaa/US07\_NEW PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubaa/PCTUS PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubaa/US08\_NEW PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubaa/US09 NEW PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubaa/US10 NEW PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubaa/US60\_NEW PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description         |
|------------|-------|-------------|--------|----|---------------------|
| 1          | 764   | 100.0       | 156    | 8  | US-08-825-486-10    |
| 2          | 764   | 100.0       | 156    | 8  | US-08-870-434-11    |
| 3          | 764   | 100.0       | 156    | 9  | US-09-372-044-10    |
| 4          | 764   | 100.0       | 156    | 9  | US-09-934-249-5     |
| 5          | 764   | 100.0       | 156    | 10 | US-09-560-150-11    |
| 6          | 764   | 100.0       | 156    | 14 | US-10-067-741-11    |
| 7          | 764   | 100.0       | 166    | 14 | US-10-106-698-5647  |
| 8          | 259   | 33.9        | 83     | 14 | US-10-106-698-6726  |
| 9          | 106.5 | 13.9        | 2971   | 14 | US-10-146-473-50    |
| 10         | 103   | 13.5        | 445    | 15 | US-10-369-493-8998  |
| 11         | 101   | 13.2        | 670    | 10 | US-09-298-523B-63   |
| 12         | 101   | 13.2        | 711    | 10 | US-09-298-523B-3    |
| 13         | 96.5  | 12.9        | 1000   | 14 | US-10-128-714-3305  |
| 14         | 97    | 12.7        | 381    | 15 | US-10-369-493-12930 |
| 15         | 96.5  | 12.6        | 449    | 15 | US-10-369-493-12408 |

|    |      |      |       |    |                     |
|----|------|------|-------|----|---------------------|
| 16 | 96.5 | 12.6 | 1259  | 14 | US-10-260-715-8     |
| 17 | 96.5 | 12.6 | 19662 | 15 | US-10-084-846A-6    |
| 18 | 96   | 12.6 | 616   | 14 | US-10-251-661-12    |
| 19 | 95.5 | 12.5 | 843   | 14 | US-10-205-823-401   |
| 20 | 94.5 | 12.4 | 1228  | 10 | US-09-917-384-1     |
| 21 | 94.5 | 12.4 | 1228  | 10 | US-09-917-383-1     |
| 22 | 94   | 12.3 | 411   | 15 | US-10-108-260A-4496 |
| 23 | 94   | 12.3 | 815   | 15 | US-10-394-322A-23   |
| 24 | 93.5 | 12.2 | 291   | 14 | US-10-029-386-32459 |
| 25 | 93.5 | 12.2 | 923   | 15 | US-10-258-106-7     |
| 26 | 93.5 | 12.2 | 2623  | 15 | US-10-374-780A-1433 |
| 27 | 93   | 12.2 | 287   | 14 | US-10-017-161-1320  |
| 28 | 93   | 12.2 | 287   | 15 | US-10-292-798-1084  |
| 29 | 92.5 | 12.1 | 113   | 9  | US-09-864-761-39507 |
| 30 | 92.5 | 12.1 | 666   | 15 | US-10-104-047-2217  |
| 31 | 92.5 | 12.1 | 802   | 9  | US-09-823-240-2     |
| 32 | 92   | 12.0 | 584   | 14 | US-10-156-761-12405 |
| 33 | 91.5 | 12.0 | 172   | 9  | US-09-764-847-551   |
| 34 | 91.5 | 12.0 | 172   | 14 | US-10-092-154-551   |
| 35 | 91.5 | 12.0 | 433   | 14 | US-10-156-761-14327 |
| 36 | 91.5 | 12.0 | 615   | 14 | US-10-109-791A-20   |
| 37 | 91   | 11.9 | 701   | 10 | US-09-298-523B-62   |
| 38 | 90.5 | 11.8 | 404   | 9  | US-09-764-864-1037  |
| 39 | 90.5 | 11.8 | 652   | 15 | US-10-104-047-3364  |
| 40 | 90.5 | 11.8 | 732   | 14 | US-10-156-761-12251 |
| 41 | 90   | 11.8 | 254   | 9  | US-09-216-393-126   |
| 42 | 90   | 11.8 | 254   | 14 | US-10-321-856-126   |
| 43 | 90   | 11.8 | 316   | 15 | US-10-116-275-116   |
| 44 | 90   | 11.8 | 377   | 14 | US-10-149-819-7     |
| 45 | 90   | 11.8 | 677   | 14 | US-10-103-313-384   |

ALIGNMENTS

RESULT 1

US-08-825-486-10  
; Sequence 10, Application US/08825486  
; Publication No. US20020016303A1  
; GENERAL INFORMATION:  
; APPLICANT: Faib, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/825,486  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-077-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE 384, App

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-825-486-10

Query Match      100.0%; Score 764; DB 8; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 60
DB 10 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 69
QY 61 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 120
DB 70 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 129
QY 121 EPNLTSPSYALDLSLTFLOQHPAAF 147
DB 130 EPNLTSPSYALDLSLTFLOQHPAAF 156

RESULT 2
US-08-870-434-11
; Sequence 11, Application US/08870434
; Publication No. US20020034736A1
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SOURCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,434
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-870-434-11

Query Match      100.0%; Score 764; DB 8; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 60
DB 10 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 69
QY 61 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 120
DB 70 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 129
QY 121 EPNLTSPSYALDLSLTFLOQHPAAF 147
DB 130 EPNLTSPSYALDLSLTFLOQHPAAF 156

; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-825-486-10

Query Match      100.0%; Score 764; DB 8; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 60
DB 10 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 69
QY 61 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 120
DB 70 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 129
QY 121 EPNLTSPSYALDLSLTFLOQHPAAF 147
DB 130 EPNLTSPSYALDLSLTFLOQHPAAF 156

RESULT 3
US-09-372-044-10
; Sequence 10, Application US/09372044A
; Patent No. US20020182603A1
; GENERAL INFORMATION:
; APPLICANT: Dean Falb et al.
; TITLE OF INVENTION: Compositions and Methods for the
; TREATMENT AND DIAGNOSIS OF Cardiovascular Disease
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-372-044-10

Query Match      100.0%; Score 764; DB 9; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 60
DB 10 TMTILQAPTAPSTIPGRRGSGPEITFDLPPEPAAAPAGRPSASGRHKRSRVLYR 69
QY 61 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 120
DB 70 VVRQLPVEENPAKLLFLITIVFCQILMAEGVPAPLPEDAPNAASLAPTVPSPVL 129
QY 121 EPNLTSPSYALDLSLTFLOQHPAAF 147
DB 130 EPNLTSPSYALDLSLTFLOQHPAAF 156

RESULT 4
US-09-934-249-5
; Sequence 5, Application US/09934249
; Patent No. US20020115081A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Richard T.
; APPLICANT: Landschulz, Katherine T.
; APPLICANT: Turi, Thomas G.
; APPLICANT: Thompson, John P.
; APPLICANT: Kennedy, Scott P.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
; CARDIOVASCULAR CONDITIONS
; FILE REFERENCE: P0738/7001/REP/KA
; CURRENT APPLICATION NUMBER: US/09/934,249
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/227,159
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 156
; TYPE: PRT

```



```
; ORGANISM: Homo Sapiens
US-09-934-249-5

Query Match      100.0%; Score 764; DB 9; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 60
   |||||
DB 10 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 69
   |||||

QY 61 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 120
   |||||
DB 70 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 129
   |||||

QY 121 EPNLTSPSYALDLSTFLQHPAAF 147
   |||||
DB 130 EPNLTSPSYALDLSTFLQHPAAF 156
   |||||

RESULT 5
US-09-560-150-11
; Sequence 11, Application US/09560150
; Publication No. US20030073076A1
; GENERAL INFORMATION:
; APPLICANT: FALB, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US/09/560,150
; PRIOR FILING DATE: 1998-07-30
; PRIOR FILING DATE: 1997-06-06
; PRIOR FILING DATE: 1997-06-06
; PRIOR FILING DATE: 1997-02-13
; PRIOR FILING DATE: 1997-02-13
; PRIOR FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq For Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-560-150-11

Query Match      100.0%; Score 764; DB 10; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.1e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 60
   |||||
DB 10 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 69
   |||||

QY 61 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 120
   |||||
DB 70 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 129
   |||||

QY 121 EPNLTSPSYALDLSTFLQHPAAF 147
   |||||
DB 130 EPNLTSPSYALDLSTFLQHPAAF 156
   |||||

RESULT 6
US-10-067-741-11
; Sequence 11, Application US/10067741
; Publication No. US20030097668A1
; GENERAL INFORMATION:
; APPLICANT: Dean A. Falb
; APPLICANT: Katherine Galvin
; APPLICANT: Michael Donovan
; APPLICANT: Dennis Huszar
US-10-067-741-11

Query Match      100.0%; Score 764; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 60
   |||||
DB 10 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 69
   |||||

QY 61 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 120
   |||||
DB 70 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 129
   |||||

QY 121 EPNLTSPSYALDLSTFLQHPAAF 147
   |||||
DB 130 EPNLTSPSYALDLSTFLQHPAAF 156
   |||||

RESULT 7
US-10-106-698-5647
; Sequence 5647, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypepti
; FILE REFERENCE: PA005PI
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5647
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5647

Query Match      100.0%; Score 764; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.3e-57;
Matches 147; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 60
   |||||
DB 10 TWIILOATPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGRHKRSRVLYPR 69
   |||||

QY 61 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 120
   |||||
DB 70 VVRQLPVEENPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPTVPSPVL 129
   |||||

QY 121 EPNLTSPSYALDLSTFLQHPAAF 147
   |||||
DB 130 EPNLTSPSYALDLSTFLQHPAAF 156
   |||||
```

Db 20 TWTLQAPTAPSTIPGRSGPEIETFDPLPEAAAPAGPSASGRHKESRRVLYPR 79  
QY 61 VVRGQLPVENPNAKELLTLTVFCQILMAEGVPAPLPPEADPNAASLAPTPVSPVL 120  
Db 80 VVRGQLPVENPNAKELLTLTVFCQILMAEGVPAPLPPEADPNAASLAPTPVSPVL 139  
QY 121 EFPNLTSEPSDYALDLSFLQHPAAF 147  
Db 140 EFPNLTSEPSDYALDLSFLQHPAAF 166  
RESULT 8  
US-10-106-698-6726  
; Sequence 6726, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
; FILE REFERENCE: PA00591  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 6726  
; LENGTH: 83  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (1)\_FEATURE  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (7)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (27)\_FEATURE  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (30)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (43)\_FEATURE  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (54)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (59)\_FEATURE  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (60)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (62)\_FEATURE  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: MISC\_FEATURE  
; LOCATION: (80)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-106-698-6726  
Query Match 33.9%; Score 259; DB 14; Length 83;  
Best Local Similarity 89.5%; Pred. NO. 7.1e-15;  
Matches 51; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
QY 91 MAEGVPAPLPPEADPNAASLAPTPVSPVLEFPNLTSEPSDYALDLSFLQHPAAF 147

Db 2 MAEGVPAPLPPEADPNAASLAPTPVSPVLEFPNLTSEPSDYALDLSFLQHPAAF 58  
RESULT 9  
US-10-146-473-50  
; Sequence 50, Application US/10146473  
; Publication No. US20030108888A1  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Matthew  
; APPLICANT: Gout, Ivan  
; APPLICANT: Stockert, Elisabeth  
; APPLICANT: Gure, Ali  
; APPLICANT: Chen, Yao-Tseng  
; APPLICANT: Old, Lloyd  
; TITLE OF INVENTION: Breast Cancer Antigens  
; FILE REFERENCE: L00451/70130(JRV)  
; CURRENT APPLICATION NUMBER: US/10/146,473  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/291,150  
; PRIOR FILING DATE: 2001-05-15  
; NUMBER OF SEQ ID NOS: 82  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 50  
; LENGTH: 2971  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-146-473-50  
Query Match 13.9%; Score 106.5; DB 14; Length 2971;  
Best Local Similarity 29.4%; Pred. NO. 3;  
Matches 37; Conservative 11; Mismatches 43; Indels 35; Gaps 5;  
QY 9 TPAPSTIPGRSGPEIETFDPLPEAAAPAGPSASGRHKESRRVLYPRVVRQLPV 68  
Db 2171 SPAREVPRP-----APR-----PRTPASAPALPA-----LVFVPSAPVFI 2209  
QY 69 EENPAKRLLELLTVFCQILMAEGVPAPLPPEADPNAASLAPTPVSPVLEFPNLTSE 128  
Db 2210 SARNP-----ITILPVHIL-----PSPPPSQIPPCSSPACTPPACTPPAHTPP 2255  
QY 129 PSDYAL 134  
Db 2256 PAQTCL 2361  
RESULT 10  
US-10-369-493-8998  
; Sequence 8998, Application US/10369493  
; Publication No. US2003023675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 8998  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Chloroflexus aurantiacus  
US-10-369-493-8998  
Query Match 13.5%; Score 103; DB 15; Length 445;  
Best Local Similarity 28.5%; Pred. NO. 0.74;  
Matches 47; Conservative 12; Mismatches 40; Indels 66; Gaps 10;



Job time : 35 secs

```
; SEQ ID NO 12930
; LENGTH: 381
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(381)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12930

Query Match          12.7%; Score 97; DB 15; Length 381;
Best Local Similarity 27.3%; Pred. No. 2;
Matches 54; Conservative 20; Mismatches 62; Indels 52; Gaps 10;

QY 1 TMTLQAPTAPSTIPGRRGSGPEIFTFDELPEPAAAPAGRPSASRGH-----RKRS 53
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 54 RRVLYPRV-----VRRQLPVEENPAKLLFL--LTIVF---COILM---AEEGVP 97
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 EVINQMSMGFERQINRAARAFNPDRATIEYLLNLSLVVFTXQLLISAFKRGIP 181
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 98 -----APLPEDAPNAASLAPTPVSPVLEPNL-----TS 127
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 ENIQOQQQSSAATTPAQA--AASGAPATSGEDEPNLFEAAQAQSGRGAGGAS 239
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 128 EPSDYALDLSTFLOQHPA 145
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 240 GGEPOSLD---FLRNHPA 254
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-10-369-493-12408
; Sequence 12408, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12408
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-369-493-12408

Query Match          12.6%; Score 96.5; DB 15; Length 449;
Best Local Similarity 27.7%; Pred. No. 2.7;
Matches 36; Conservative 10; Mismatches 47; Indels 37; Gaps 6;

QY 8 PTPAPSTIPGRRGSGPEIFTFDELPEPAAAPAGRPSASRGHKKRS-----RVLYPRVV 62
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 34 PPGPGPGPGFH--GPHSYSPHPPQPGVAPSGPPGPGYQPGYQYNAQGYPP-- 88
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 63 RROLPVEERNPAKLLFLTLTVFCOILMAEKGVPAP-----LPEPAPNAAS 110
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 89 -----EPHSPQFPFQ-----HPQHFQPPGPPGPPGPHGYGQFGAPPMFAPSMPS 135
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 111 L--AptPVSP 118
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 136 LGYAPGQVAP 145
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: February 25, 2004, 05:24:45 ; Search time 20 Seconds  
(without alignments)  
707.008 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_10\_156  
Perfect score: 764  
Sequence: 1 TWIILQAPTPAPSTIPGRPR.....EPDVALDLSTFLQHPAAF 147

Scoring table: ELOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: PIR1:\*  
2: PIR2:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | ID       | Description        |
|------------|-------|-------|--------|----------|--------------------|
| 1          | 760   | 99.5  | 156    | JC5537   | differentiation-de |
| 2          | 501   | 65.6  | 153    | S33363   | gly96 protein - mo |
| 3          | 107   | 14.0  | 496    | T17908   | proline/lysine-ric |
| 4          | 106.5 | 13.9  | 581    | T22341   | hypothetical prote |
| 5          | 106   | 13.9  | 1029   | T30351   | mucin-like protein |
| 6          | 105.5 | 13.8  | 283    | S13383   | hydroxyproline-ric |
| 7          | 105   | 13.7  | 549    | T17825   | proline-rich prote |
| 8          | 104.5 | 13.7  | 535    | S76953   | protein kinase (EC |
| 9          | 104.5 | 13.7  | 552    | T08148   | proline-rich myros |
| 10         | 104.5 | 13.7  | 3149   | 1 Q08E8  | BRF1 protein - hu  |
| 11         | 104   | 13.6  | 351    | S50754   | hypothetical prote |
| 12         | 102.5 | 13.4  | 288    | T17737   | proline-rich prote |
| 13         | 102.5 | 13.4  | 901    | A49227   | sialidase - Actino |
| 14         | 102.5 | 13.4  | 3869   | 2 A48205 | Al1-1 protein +GRE |
| 15         | 102   | 13.4  | 708    | T09519   | hypothetical prote |
| 16         | 101   | 13.2  | 339    | 2 D96711 | proline-rich prote |
| 17         | 97.5  | 12.8  | 225    | 2 T17815 | proline-rich prote |
| 18         | 97.5  | 12.8  | 1323   | 2 T00037 | hypothetical prote |
| 19         | 97.5  | 12.8  | 3938   | 2 T42761 | Bassoon protein -  |
| 20         | 97    | 12.7  | 473    | 2 S50755 | hypothetical prote |
| 21         | 96    | 12.6  | 185    | 2 T00519 | proline-rich prote |
| 22         | 96    | 12.6  | 461    | 2 T10741 | extensin-like prot |
| 23         | 95.5  | 12.5  | 211    | 2 B89716 | protein F45B8.3 (i |
| 24         | 95.5  | 12.5  | 241    | 2 T22216 | hypothetical prote |
| 25         | 95.5  | 12.5  | 524    | 2 A75588 | probable protein k |
| 26         | 95.5  | 12.5  | 753    | 2 JQ0532 | OP protein - Keme  |
| 27         | 95    | 12.4  | 202    | 2 JQ0964 | hydroxyproline-ric |
| 28         | 95    | 12.4  | 416    | 1 SKXLAG | dermal gland prote |
| 29         | 95    | 12.4  | 801    | 2 T29018 | hypothetical prote |

30 94 12.3 815 2 B56708 extracellular sign  
31 94 12.3 1290 2 T00018 period protein hom  
32 93.5 12.2 316 2 T31880 hypothetical prote  
33 93.5 12.2 537 2 A46611 myosin-binding pro  
34 93.5 12.2 555 2 T30349 structural protein  
35 93.5 12.2 654 2 T34960 cell division prot  
36 93.5 12.2 1402 2 T46707 translation initia  
37 93 12.2 135 2 T49996 AtAGP4 - Arabidops  
38 93 12.2 684 2 T25603 hypothetical prote  
39 93 12.2 1001 2 T16419 hypothetical prote  
40 93 12.2 1165 2 T16420 hypothetical prote  
41 93 12.2 1801 2 T26774 Bassoon protein -  
42 93 12.2 3942 2 T42730 Bassoon protein -  
43 92.5 12.1 275 2 T24593 hypothetical prote  
44 92.5 12.1 279 2 T10361 hypothetical prote  
45 92.5 12.1 605 1 Q08E29 BRF1 protein - hu

ALIGNMENTS

RESULT 1

JC5537  
differentiation-dependent protein DIF-2 - human  
C:Species: Homo sapiens (man)  
C:Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999  
C:Accession: JC5537  
R:Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.  
Biochem. Biophys. Res. Commun. 235, 4-9, 1997  
A:Title: Identification and characterization of a novel monocyte/macrophage differentie  
A:Reference number: JC5537; MUID:97339426; PMID:9196025  
A:Accession: JC5537  
A:Molecule type: mRNA  
A:Residues: 1-156 <EIE>  
A:Experimental source: monocyte  
A>Note: the authors translated the codon CCG for residue 106 as Arg

Query Match 99.5%; Score 760; DB 2; Length 156;  
Best Local Similarity 99.3%; Pred. No. 9.1e-55;  
Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TWIILQAPTPAPSTIPGRSGPFIITFDLPPEAAAPAGPSASGRHXRSRVLYPR 60  
Db 10 TWIILQAPTPAPSTIPGRSGPFIITFDLPPEAAAPAGPSASGRHXRSRVLYPR 69  
Qy 61 VVRQQLPVEENPAKRLIFLLITVFCQILMAEGVPAPLPEDAPNAASLAPTFVSPVL 120  
Db 70 VVRQQLPVEENPAKRLIFLLITVFCQILMAEGVPAPLPEDAPNAASLAPTFVSPVL 129  
Qy 121 EPNLTSEPSDYALDLSTFLQHPAAF 147  
Db 130 EPNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2

S33363  
gly96 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997  
C:Accession: S33363  
R:Charles, C.H.; Yoon, J.K.; Simske, J.S.; Lau, L.F.  
Oncogene 9, 797-801, 1993  
A:Title: Genomic structure, cDNA sequence, and expression of gly96, a growth factor-inc  
A:Reference number: S33363; MUID:93173526; PMID:8437864  
A:Accession: S33363  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-153 <CHA>  
A:Cross-references: EMBL:X67644  
C:Genetics:  
A:Introns: 70/3  
C:Keywords: transmembrane protein

```

Rest Local Similarity 25.8%; Pred. No 0.26;
Matches 41; Conservative 9; Mismatches 54; Indels 55; Gaps 6;

QY      7  ATPAPSTTIGPRGGPIETFDLPPEAAAP-----ACGPGASGCHKKSR  54
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     420  APPPAPEVPA-----PAPAPEAFAVPSADAGYAAAAAPAGGSGVPKKR  466

QY     55  RV-----LYRVRVRQLPVEPNPAKLLFLITVFCILMAEGVPALP-----  101
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     467  RVARYAAGEAATAAEAPAPAPAPA-----PEAPAPEPAPAPAPAPE  516

QY    102  --PEDAPNAASLAPTV-----SPVLFPNLTSEPDYA  133
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    517  AAFAAAPDAAPAFAPVPEVAPAPAPAPAPAPESDAGSGYS  555

```

RESULT 5  
T30351

C:Species: Lymnathia dispar nuclear polyhedrosis virus, LdMNPV  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change  
 C:Accession: T30351  
 R:Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rolfs  
 Virology 253, 1-34, 1999  
 A:Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymnathia  
 A:Reference number: Z20836; MUID:99124785; PMID:9887315  
 A:Accession: T30351  
 A:Status: preliminary; translated from GB/EWEL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1029 <KUZ>  
 A:Cross-references: EWEL:AF034810; PIDN:AACT070189.1

```

Query March      13.9%; Score 106; DB 2; Length 1029;
Best Local Similarity 29.5%; Pred. No. 0.5;
Matches 51; Conservative 6; Mismatches 52; Indels 64; Gaps 8

QY      7  APTAPSPITGPRQSGP-EITF-----DPLPEPAAAPAGPFSAGSRGHRKSRVLY 58
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      637  ADEPAPETVEP-----SAPVDYFTINGAETAPAPAPAPAPETPKETSEPALG 693

QY      59  PAVRRQCPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTTVP 118
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      694  P-VEPALEPVE-----PALEPVEPAPGCEPALERPGE 725

QY      119  VLEPPTLSE-----PSDYALDI-----STF-----LQHPAA 146
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      726  ALEPGESTAPAGELAVERPAPAPDITSDSAVESTFGHKSTVASSLAARAA 778

```

## C;Species

hydroxyproline-rich glycoprotein - sorghum  
 C:Species: Sorghum bicolor (sorghum)  
 C:date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 24-Sep-1999  
 C:Accession: S13383; S14449  
 R:Raz, R.; Cretin, C.; Puigdomenech, P.; Martinez-Izquierdo, J.A.  
 Plant Mol. Biol. 16, 365-367, 1991  
 A:title: The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum vulgare.  
 A:Reference number: S13383; MUID:91370882; PMID:1933107  
 A:Accession: S13383  
 A:Molecule type: DNA  
 A:Molecule type: DNA  
 A:Residues: 1-283 <RAZI>  
 A:Cross-references: EMBL:X56010  
 A:Note: the source is designated as Sorghum vulgare  
 R:Raz, R.  
 submitted to the EMBL Data Library, October 1990  
 A:Reference number: S14449  
 A:Accession: S14449  
 A:Molecule type: DNA  
 A:Residues: 1-238, 'Y', 240, 'QAH', 244-283 <RAZ2>  
 A:Cross-references: EMBL:X56010; NID:g21626; PIDN:CAA39485.1; PID:g21627  
 C:Superfamily: hydroxyproline-rich glycoprotein  
 C:Keywords: glycoprotein; hydroxyproline

Query Match 13.8%; Score 105.5; DB 2; Length 283;  
Best Local Similarity 23.8%; Pred. No. 0.15;  
Matches 41; Conservative 9; Mismatches 81; Indels 41; Gaps 4;

QY 1 TMTLQAPTAPST-----IPGRGSGPEIFTFDPLPSPAAPAGR 42  
DB 73 TTPSEKTPPPATPKPTPTTPSPKPKSVVPPPKASTPTPTSPKP-PATKPTY 131  
QY 43 PSASRGHKKSRVLY-----PRVVRQLPVEBPNFAKELFLLLTIVFCQILMA 92  
DB 132 PTKPKPAKPPTPPVTPSPKPKPTPTTPSPKPKPTPTTPSPKPKPTPTTPSPK 179  
QY 93 ESGVAPLIPEDAPNAASLAPTPSPVLPEPNITSEPSDYALDSTLQHP 144  
DB 180 PTHTPSPKPKPTPTTPPVTPSPKPKPTPTTPSPKPKPTPTTPSPKPKPTPTTP 231

RESULT 7  
T17525  
proline-rich protein A35L - Chlorella virus PBCV-1  
C/Species: Chlorella virus PBCV-1  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C/Accession: T17525  
R/Graves, M.V.; Van Baten, J.L.  
submitted to the EMBL Data Library, May 1999  
A/Reference number: Z18806  
C/Accession: T17525  
A/Status: preliminary; translated from GB/EMBL/DDSV  
A/Molecule type: DNA  
A/Residues: 1-549 <GRA>  
A/Cross-references: EMBL:U42580; NID:G4028896; PIDN:AA096403.1  
A/Experimental source: specific host Chlorella strain NC64A  
C/Genetics:  
A/Gene: A35L

Query Match 13.7%; Score 105; DB 2; Length 549;  
Best Local Similarity 27.8%; Pred. No. 0.32;  
Matches 35; Conservative 11; Mismatches 36; Indels 44; Gaps 6;

QY 7 APTAPSTIPGFRGGGRIPTFDPLPEPAAPAGPSASRGRKRRLVPRVVRQL 66  
DB 445 APKPAKPAPKPAKPAPK-----PAKPAKPAPKPA-----PKPAKPA 485  
QY 67 PVEBPNPAKRLFLLLTIVFCQILMAESGVPAPLP-PEDAPNAASLAPTPV-SVLPEFN 124  
DB 486 PKPAKPAPK-----PAKPAKPAPKAPK-----APKPAKPAPKPAV 522

QY 125 LASEPS 130  
DB 523 ITSKVS 528

RESULT 8  
S76953  
protein kinase (BC 2.7.1.-), 58K - Synecocystis sp. (strain PCC 6803)  
N/Alternate names: protein slr0599  
C/Species: Synecocystis sp.  
A/Variety: PCC 6803  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 27-Oct-2003  
R/Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimoto, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis  
s.  
A/Reference number: S74322; MUID:97061201; PMID:8905231  
A/Accession: S76953  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-535 <KAN>  
A/Cross-references: EMBL:D90917; GB:AB001339; NID:G1653836; PIDN:BAAL8865.1; PID:G165395  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C/Superfamily: protein kinase, 58 kDa type; protein kinase homology  
C/Keywords: phosphotransferase; protein kinase  
F;10-274/Domain: protein kinase homology <KIN>

Query Match 13.7%; Score 104.5; DB 1; Length 535;  
Best Local Similarity 25.6%; Pred. No. 0.34;  
Matches 32; Conservative 9; Mismatches 43; Indels 41; Gaps 4;

QY 7 APTAPSTIPGFRGGGRIPTFDPLPEPAAPAGPSASRGRKRRLVPRV 60

DB 411 SPEPTSPSPFETTSPTEDTITMEPEPSLDPEPAIPEPKPSPS----- 456

QY 61 VVRQLPVEBPNPAKRLFLLLTIVFCQILMAESGVPAPLP-PEDAPNAASLAPTPVSV 119

DB 457 -----PTISQPSPTISIPV-----TPAPVPSPSPPTPKPTVPPQISPT 496

QY 120 LEFFN 124

DB 497 PQPSN 501

## RESULT 9

T08148

proline-rich myrosinase-binding protein homolog - rape (fragment)

N/Alternate names: myrosinase-binding protein related protein

C/Species: Brassica napus (rape)

C/Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 08-Oct-1999

C/Accession: T08148

R/Taipalensuu, J.; Falk, A.; Ek, B.; Rask, L.

Eur. J. Biochem. 243, 605-611, 1997

A/Title: Myrosinase-binding proteins are derived from a large wound-inducible and repe-

A/Reference number: Z16379; MUID:97210758; PMID:9057822

A/Accession: T08148

A/Status: preliminary; translated from GB/EMBL/DDBV

A/Molecule type: mRNA

A/Residues: 1-552 <TAI>

A/Cross-references: EMBL:U59446; NID:G1655829; PIDN:AA08051.1; PID:G1655830

A/Experimental source: cv. 20516 of Svalof's Karat; immature seed, 30 days after pollina-

Query Match 13.7%; Score 104.5; DB 2; Length 552;

Best Local Similarity 29.4%; Pred. No. 0.36;

Matches 40; Conservative 6; Mismatches 37; Indels 53; Gaps 7;

QY 7 APTAPSTIPGFRGGGRIPTFDPLPEPAAP-----AG-RPSASRGRKR 54

DB 301 APSAPGAPAPAGPSHPA-----PAPAPAPQGPRPAPAGPAPGPRSPAPGPR-- 353

QY 55 RVLYPRVVRQLPVEBPNPAKRLFLLLTIVFCQILMAESGVPAPLP---PEDAPNA--- 108

DB 354 ----PAPAPAPAGQGPRPA-----PGAPQGPFHPAPAPAPAG 387

QY 109 --ASLAPTPVSVLEP 122

DB 388 TSATPAPATTTKGP 403

## RESULT 10

Q2B83

BPUF1 protein - human herpesvirus 4 (strain B95-8)

C/Species: human herpesvirus 4, Epstein-Barr virus

C/Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 16-Jul-1999

C/Accession: G93065; A03747; S32993

R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr viru

A/Reference number: A93065; MUID:85035713; PMID:6092825

A/Accession: G93065

A/Molecule type: DNA

A/Residues: 1-3149 <BAN>

A/Cross-references: EMBL:V01555; NID:G59074; PIDN:CA244939.1; PID:gl334853

R/Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.;

Nature 310, 207-211, 1984

A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.





Db 1.132 TSVKGLEFAQAAAPRPEAPAKRSSEPPKPKVEKSEGGAPAPAPBPBKQVSAP 1191  
QY 48 GHRKSRVLRVPRVRQRQPVPEENPAKRLLLFLLITVFCQILMASEGVA-----PLP 101  
Db 1.192 ASKSKSQVSPAAV---VPPQFSTAP-----OKKAPKAVPSEPKKQPPP 1236  
QY 102 PEDAPNAA---SLAPTPVSPVLEPNTLSTSPSD 131  
Db 1237 PEGPSEQSKQKVALEPSPVQK-----KPKD 1263

RESULT 15  
D96711  
hypothetical protein F24J5.8 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: D96711  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.N.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 836-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A96141; MUID:21016719; PMID:11130712  
A:Accession: D96711  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-708 <STO>  
A:Cross-references: GB:AB005173; NID:G5734709; PIDN:AA349974.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F24J5.8  
A:Map position: 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:21:10 ; Search time 17 Seconds  
(without alignments)  
450.254 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_10\_156

Perfect score: 764

Sequence: 1 TWTLQAPTPAPSTIPGPRR.....EPSDVALDLSTFLQHPAAF 147

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 764   | 100.0       | 156    | 1     | TEXT1_HUMAN |
| 2          | 757   | 99.1        | 156    | 1     | TEXT1_PANTR |
| 3          | 501   | 65.6        | 153    | 1     | TEXT1_MOUSE |
| 4          | 105.5 | 13.8        | 283    | 1     | EXTN_SORBI  |
| 5          | 104.5 | 13.7        | 535    | 1     | SPKC_SINY3  |
| 6          | 104.5 | 13.7        | 3149   | 1     | TEGU_EBV    |
| 7          | 103.5 | 13.5        | 426    | 1     | ASD4_NEUCR  |
| 8          | 102.5 | 13.4        | 381    | 1     | APA_MYCAV   |
| 9          | 102.5 | 13.4        | 3866   | 1     | HRX_MOUSE   |
| 10         | 101.5 | 13.3        | 268    | 1     | NO20_MEDTR  |
| 11         | 98    | 12.8        | 864    | 1     | WS14_MOUSE  |
| 12         | 97    | 12.7        | 852    | 1     | WS14_HUMAN  |
| 13         | 96    | 12.6        | 616    | 1     | VGF_HUMAN   |
| 14         | 95.5  | 12.5        | 806    | 1     | MK07_MOUSE  |
| 15         | 95.5  | 12.5        | 964    | 1     | MRTA_MOUSE  |
| 16         | 95    | 12.4        | 439    | 1     | XP2_XENIA   |
| 17         | 94    | 12.3        | 815    | 1     | MK07_HUMAN  |
| 18         | 94    | 12.3        | 1290   | 1     | PERL_HUMAN  |
| 19         | 93.5  | 12.2        | 537    | 1     | MYPH_CHICK  |
| 20         | 93.5  | 12.2        | 1402   | 1     | IP4G_RABIT  |
| 21         | 93    | 12.2        | 1647   | 1     | POEL_CABEL  |
| 22         | 92.5  | 12.1        | 279    | 1     | Y091_NPVOP  |
| 23         | 92.5  | 12.1        | 605    | 1     | BRLT_EBV    |
| 24         | 92.5  | 12.1        | 625    | 1     | NIFA_AZOB   |
| 25         | 92.5  | 12.1        | 802    | 1     | ENAH_MOUSE  |
| 26         | 92.5  | 12.1        | 837    | 1     | GCL2_HUMAN  |
| 27         | 92    | 12.0        | 3969   | 1     | HRX_HUMAN   |
| 28         | 91.5  | 12.0        | 510    | 1     | PERM_HUMAN  |
| 29         | 91.5  | 12.0        | 615    | 1     | MUTL_ECOLI  |
| 30         | 91.5  | 12.0        | 1133   | 1     | SREI_CRIGR  |
| 31         | 91.5  | 12.0        | 1595   | 1     | SOS_DROME   |
| 32         | 91    | 11.9        | 3511   | 1     | MY15_MOUSE  |
| 33         | 90.5  | 11.8        | 3164   | 1     | TEGU_HSV11  |

|    |      |      |      |   |            |
|----|------|------|------|---|------------|
| 34 | 90   | 11.8 | 316  | 1 | CDNC_HUMAN |
| 35 | 90   | 11.8 | 449  | 1 | APG_ERANA  |
| 36 | 90   | 11.8 | 914  | 1 | GNDS_HUMAN |
| 37 | 90   | 11.8 | 959  | 1 | FTKI_RALSO |
| 38 | 90   | 11.8 | 1509 | 1 | GSRI_HUMAN |
| 39 | 89   | 11.6 | 534  | 1 | APG_ARATH  |
| 40 | 89   | 11.6 | 817  | 1 | VRPI_YEAST |
| 41 | 89   | 11.6 | 964  | 1 | IF2_CHEVO  |
| 42 | 89   | 11.6 | 1046 | 1 | IF2_STRAW  |
| 43 | 89   | 11.6 | 1291 | 1 | PERI_MOUSE |
| 44 | 89   | 11.6 | 5262 | 1 | MLL2_HUMAN |
| 45 | 88.5 | 11.6 | 615  | 1 | MUTL_ECOLI |

## ALIGNMENTS

|             |   |           |      |         |  |
|-------------|---|-----------|------|---------|--|
| RESULT 1    |   |           |      |         |  |
| TEXT1_HUMAN |   |           |      |         |  |
| ID          | TEXT1_HUMAN   | STANDARD; | PRT; | 156 AA. |  |
| AC          | P46695; Q32691; Q93044;   |           |      |         |  |
| DT          | 01-NOV-1995 (Rel. 32, Created)  |           |      |         |  |
| DT          | 15-JUL-1998 (Rel. 36, Last sequence update)   |           |      |         |  |
| DT          | 15-MAR-2004 (Rel. 43, Last annotation update)   |           |      |         |  |
| DE          | Radiation-inducible immediate-early gene IX-1 (Immediate early protein GLY96) (Immediate early response 3 protein) (PACAP-responsive gene 1 protein) (PRGI protein) (Differentiation-dependent gene 2 protein) (DIF-2 protein).   |           |      |         |  |
| DE          | gene 1 protein (PRGI protein) (Differentiation-dependent gene 2 protein) (DIF-2 protein).   |           |      |         |  |
| GN          | IER3 OR IEX1 OR PRG1 OR DIF2.   |           |      |         |  |
| OS          | Homo sapiens (Human).   |           |      |         |  |
| OC          | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |           |      |         |  |
| OX          | NCBI_TaxID=9606;  |           |      |         |  |
| RN          | [1]   |           |      |         |  |
| RP          | SEQUENCE FROM N.A.  |           |      |         |  |
| RC          | TISSUE=Placenta;  |           |      |         |  |
| EX          | MEDLINE=96181295; PubMed=8603392;   |           |      |         |  |
| RA          | Kondratyev A.D., Chung K.-N., Jung M.O.;  |           |      |         |  |
| RT          | "Identification and characterization of a radiation-inducible glycosylated human early-response gene."  |           |      |         |  |
| RL          | Cancer Res. 56:1499-1502(1996).   |           |      |         |  |
| RN          | [2]   |           |      |         |  |
| RP          | SEQUENCE FROM N.A.  |           |      |         |  |
| EX          | MEDLINE=96221139; PubMed=8653710;   |           |      |         |  |
| RA          | Schaefer H., Trautold A., Siegel E.G., Folsch U.R., Schmidt W.E.;   |           |      |         |  |
| RT          | "PRGI: a novel early-response gene transcriptionally induced by pituitary adenylate cyclase activating polypeptide in a pancreatic carcinoma cell line."  |           |      |         |  |
| RL          | Cancer Res. 56:2641-2641(1996).   |           |      |         |  |
| RN          | [3]   |           |      |         |  |
| RP          | SEQUENCE FROM N.A.  |           |      |         |  |
| EX          | MEDLINE=9739426; PubMed=9196025;  |           |      |         |  |
| RA          | Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;   |           |      |         |  |
| RT          | "Identification and characterization of a novel monocyte/macrophage differentiation-dependent gene that is responsive to lipopolysaccharide, ceramide, and lysophosphatidylcholine."  |           |      |         |  |
| RL          | Biochem. Biophys. Res. Commun. 235:4-9(1997).   |           |      |         |  |
| RN          | [4]   |           |      |         |  |
| RP          | SEQUENCE FROM N.A.  |           |      |         |  |
| EX          | Shina S., Tamiya G., Oka A., Inoko H.;  |           |      |         |  |
| RT          | "Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region."  |           |      |         |  |
| RL          | Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.   |           |      |         |  |
| RN          | [5]   |           |      |         |  |
| RP          | SEQUENCE FROM N.A.  |           |      |         |  |
| RC          | TISSUE=Cervix, and Skin;  |           |      |         |  |
| EX          | MEDLINE=22389257; PubMed=12477932;  |           |      |         |  |
| RA          | Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., |           |      |         |  |

DB protein GLY96) (Immediate early response 3 protein).  
 GN IER3 OR IEX1  
 OS Pan troglodytes (Chimpanzee).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
 NCBI\_TaxID=9598;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RP MEGLINE=22709134; PubMed=12799463;  
 RX Anzai T., Shina T., Kimura N., Yanagiya K., Kohara S., Shigenari A.,  
 RA Yamagata T., Kulski J.K., Naruse T.K., Fujimori Y., Fukuzumi Y.,  
 RA Yamazaki M., Tashiro H., Iwamoto C., Umebara Y., Imanishi T.,  
 RA Meyer A., Ikeo K., Gotohori T., Bahram S., Inoko H.;  
 RT "Comparative sequencing of human and chimpanzee MHC class I regions  
 RT unveils insertions/deletions as the major path to genomic  
 RT divergence".  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7708-7713 (2003).  
 CC -!- SUBCELLULAR LOCATION: type II membrane protein (Potential).  
 CC -!- PTM: Glycosylated (By similarity).  
 CC -!- SIMILARITY: Belongs to the IER3 family.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AB100C84; BAC78173.1; -  
 KW Glycoprotein; Transmembrane; Signal-anchor.  
 FT DOMAIN 1 82  
 FT TRANSMEM 83 99  
 FT  
 FT FT  
 FT DOMAIN 100 156  
 FT CARBOHYD 133 133  
 FT SEQUENCE 156 AA; 16929 XM; 83C067CDCAC09650 CRC64;  
 SQ  
 -----  
 Query Match 99.1%; Score 757; DB 1; Length 156;  
 Best Local Similarity 99.3%; Pred. No. 5.8e-54;  
 Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 -----  
 QY 1 TWTLQATPPAPSTIPGRRSGSGPEITFPDLPEAAPAGPSASGRHKRSRVLYPR 60  
 DB 10 TWTLQATPPAPSTIPGRRSGSGPEITFPDLPEAAPAGPSASGRHKRSRVLYPR 69  
 QY 61 VVRQLPVEEENPAKRLFLLLTVFCQILMAEEGVPAIPPEPDAPNAASLAPTPVSPVL 120  
 DB 70 VVRQLPVEEENPAKRLFLLLTVFCQILMAEEGVPAIPPEPDAPNAASLAPTPVSPVL 129  
 QY 121 EPNLTSPESDYALDLSTLQOHPAAF 147  
 DB 130 EPNLTSPESDYALDLSTLQOHPAAF 156  
 -----  
 RESULT 3  
 IEX1 MOUSE  
 ID IEX1 MOUSE STANDARD; PRT; 153 AA.  
 AC P46694;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Radiation-inducible immediate-early gene IEX-1 (Immediate early  
 DE protein GLY96) (Immediate early response 3 protein).  
 GN IER3 OR IEX1 OR GLY96.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RX SEQUENCE FROM N.A.  
 RP STRAIN=BALE/c;  
 RX MEDLINE=93173526; PubMed=8437864;

RA Charles C.H., Yoon J.K., Simske J.S., Lau L.F.;  
 RT "Genomic structure, cDNA sequence, and expression of gly96, a growth  
 RT factor-inducible immediate-early gene encoding a short-lived  
 RT glycosylated protein.";  
 RL Oncogene 8:797-801(1993).  
 CC -!- FUNCTION: Not known; expressed during the G0-G1 transition of the  
 CC cell cycle.  
 CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).  
 CC -!- TISSUE SPECIFICITY: Expressed predominantly in the lung, testes  
 CC and the uterus.  
 CC -!- INDUCTION: By serum growth factors and TPA.  
 CC -!- PTM: Glycosylated.  
 CC -!- SIMILARITY: Belongs to the IER3 family.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: X67644; ; NOT ANNOTATED\_CDS.  
 CC PIR: S33363; S33363.  
 CC MG: MGI:104814; Ier3.  
 KW Glycoprotein; Transmembrane; Signal-anchor.  
 FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 86 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT (POTENTIAL).  
 FT DOMAIN 103 153 EXTRACELLULAR (POTENTIAL).  
 FT CARBOHYD 137 137 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 153 AA; 16875 MW; 16875 MW; F4429E3B0120D11 CRC64;  
 Query Match 65.6%; Score 501; DB 1; Length 153;  
 Best Local Similarity 71.2%; Pred. No. 1.5e-33;  
 Matches 104; Conservative 8; Mismatches 26; Indels 8; Gaps 2;  
 QY 1 TWTLQATPAPSTIPGRRGSGPEIFTFDPLPEPAAGPAPGASRCHGRKRRVLYPR 60.  
 DB 10 TWTLGRAPSPAPSGPRLRGSGPEIFTFDPLPERAVWSTALNTSRGHRKRRVLYPR 69  
 QY 61 VVRQQLPVEENPAKRLIFLLITIVFCOILMAEGVPAPLPEDAPNAASLAPTVS--- 117  
 DB 70 VVRQQLPVEENPAKRLIFLLITIVFCOILMAEGVSPQLAPEDATSA--VTPEIPISAFI 127  
 QY 118 ---PVLEPNTSPSPDYALDLSTEL 140  
 DB 128 TAPVLEPINLTSSSDYALDLKAEI 153  
 RESULT 4  
 EXTN\_SORBI STANDARD; PRT; 283 AA.  
 ID AC P24152;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Extensin precursor (Proline-rich glycoprotein).  
 GN HRGP.  
 OS Sorghum bicolor (Sorghum) (Sorghum vulgare).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC PACCAD clade; Panicoideae; Andropogoneae; Sorghum.  
 OX NCBI\_TaxID=4558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Leaf;  
 RX MEDLINE=91370882; PubMed=1893107;  
 RA Raz R., Cretin C., Puigdomenech P., Martinez-Izquierdo J.A.;  
 RT "The sequence of a hydroxyproline-rich glycoprotein gene from Sorghum  
 RT vulgare.";  
 RL Plant Mol Biol. 16:365-367(1991).  
 CC -!- FUNCTION: Structural component in primary cell wall.

CC -!- PTM: EXTENSINS CONTAIN A CHARACTERISTIC REPEAT OF THE PENTAPEPTIDE  
 CC SER-PRO(4). THE PROLINE RESIDUE IS HYDROXYLATED AND THEN  
 CC GLYCOSYLATED.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL: X56010; CNA39485.1; --  
 CC PIR: S13383; S13383.  
 CC InterPro: IPR003882; Pistil\_extensin.  
 CC PRINTS: PR01218; PSTEXTENSIN.  
 KW Repeat; Cell wall; Glycoprotein; Signal; Structural protein;  
 KW Hydroxylation.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 283 EXTENSIN.  
 SQ SEQUENCE 283 AA; 29593 MW; 8D7FCD0DA8ED2D90 CRC64;  
 Query Match 13.8%; Score 105.5; DB 1; Length 283;  
 Best Local Similarity 23.8%; Pred. No. 0.097;  
 Matches 41; Conservative 9; Mismatches 81; Indels 41; Gaps 4;  
 QY 1 TWTLQATPAPST-----IPGRRGSGPEIFTFDPLPEPAAGPAPG 42  
 DB 73 TYTPSPKPTPPATPKPTPTPTTPSPKSPVYPPPKASTPTPTPSKPK-PATKPTPT 131  
 QY 43 PSASRGHRGSGPRVLY-----PRVRRQLPVEENPAKRLIFLLITIVFCOILMA 92  
 DB 132 PTKPKPATKPTPTPTPTTPSPKPVTKPTPKPTPTPTPTPTPTPTPTPTPTPTPT 179  
 QY 93 EGVGPAPLPEDAPNAASLAPTVSPVLEPNTSPSPDYALDLSTFLQOHP 144  
 DB 180 PTHTPSPKPTSKPTPTPTPTTPSPKPKSPPTPTPTPTPKPATKPTPTPTPTPTPT 231  
 RESULT 5  
 SPKC\_SVNY3 STANDARD; PRT; 535 AA.  
 ID AC P74745;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Serine/threonine-protein kinase C (EC 2.7.1.37).  
 GN SPKC OR SLR0599.  
 OS Synchocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.  
 OX NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
 RX MEDLINE=22158630; PubMed=12168951;  
 RA Kamei A., Yiasa T., Gerg X., Ikeuchi M.;  
 RT "Biochemical examination of the potential eukaryotic-type protein  
 RT kinase genes in the complete genome of the unicellular Cyanobacterium  
 RT synchocystis sp. PCC 6803.";  
 RL DNA Res. 9:71-78(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;  
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,  
 RA Shimizu S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.;  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT Synchocystis sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.





DR SMART; SM00249; PHD; 4.  
 DR SMART; SM00508; Post-SET; 1.  
 DR SMART; SM00317; SET; 1.  
 DR PROSITE; PS0014; BROMODOMAIN 2; 1.  
 DR PROSITE; PS0068; POST-SET; 1.  
 DR PROSITE; PS0280; SET; 1.  
 DR PROSITE; PS01359; ZF PHD 1; 3.  
 DR PROSITE; PS0016; ZF PHD 2; 3.  
 KW DNA-binding; Nuclear protein; Zinc-finger; Metal-binding;  
 KW Transcription regulation; Bromodomain; Alternative splicing;  
 KW Polymorphism.  
 FT NON\_TER 1 1  
 FT DNA\_BIND 67 78 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 115 125 A.T HOOK (BY SIMILARITY).  
 FT DNA\_BIND 199 207 A.T HOOK (BY SIMILARITY).  
 FT ZN\_FING 1044 1091 CXXC-TYPE.  
 FT ZN\_FING 1330 1381 PHD-TYPE 1.  
 FT ZN\_FING 1383 1432 PHD-TYPE 2.  
 FT ZN\_FING 1465 1529 PHD-TYPE 3.  
 FT DOMAIN 1605 1650 BROMODOMAIN (DIVERGENT).  
 FT DOMAIN 3725 3846 SET.  
 FT DOMAIN 3850 3866 POST-SET.  
 FT DOMAIN 35 41 POLY-GLY.  
 FT DOMAIN 459 469 PRO-RICH.  
 FT DOMAIN 1231 1238 POLY-PRO.  
 FT DOMAIN 3533 3536 POLY-GLU.  
 FT DOMAIN 3693 3697 POLY-GLU.  
 FT VARSPLIC 1503 1505 Missing (in isoform 2).  
 FT VARIANT 1497 1497 X -> T.  
 FT SEQUENCE 3866 AA; 420976 MW; ADFCS5E14E806FID CRC64;  
 Query Match 13.4%; Score 102.5; DB 1; Length 3866;  
 Best Local Similarity 26.8%; Pred. No. 2.4;  
 Matches 41; Conservative 19; Mismatches 48; Indels 45; Gaps 8;  
 QY 3 TILQAP-TPAPSTIPGRGSGPEITFDPLP-----PAAAGAPPS-ASR 47  
 DB 1129 TSVKPLEPAQKAAPPREPAPKSGSEPPRPVVEKSEGGAPAPAPAPKQVSAP 1188  
 QY 48 GHEKRRVLYVRVRLQVPEPNPAKLLFLLITVFCQILMAEGVPA-----PLP 101  
 DB 1189 ASKSKSQVQDAV---VPPQPTAP-----QKEAPKAVSEPKKQPP 1233  
 QY 102 PEDAPNA---SLAPTPVSVLPENLTSEPSD 131  
 DB 1234 PEPGQSQKKVAPLPSIPVKQ-----KPKD 1260  
 RESULT 10  
 NO20 MEDTR  
 ID NO20 MEDTR STANDARD; PRT; 268 AA.  
 AC P93329;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Early nodulin 20 precursor (N-20).  
 GN ENOD20.  
 OS Medicago truncatula (Barrel medic).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eursids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.  
 OX NCBI\_TaxID=3880;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. JewelLong;  
 RX MEDLINE=96187258; PubMed=9526510;  
 RA Greene E.A., Eard M., Dedieu A., Barker D.G.;  
 RT "NENOD16 and 20 are members of a family of phytoecyanin-related early  
 nodulins.";  
 RL Plant Mol. Biol. 36:775-783 (1998).  
 CC -!- SIMILARITY: Contains 1 plastocyanin-like domain.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; X99467; CAA67830.1; --  
 DR InterPro; IPR008972; Cupredoxin.  
 DR InterPro; IPR003245; Plicyanin like.  
 DR Pfam; PF02288; Cu\_Bind\_Like; 1.  
 DR ProDom; PD003122; Plicyanin like; 1.  
 FT SIGNAL 1 22 BY SIMILARITY.  
 FT CHAIN 23 268 EARLY NODULIN 20.  
 FT DOMAIN 23 ? PLASTOCYANIN-LIKE.  
 FT DOMAIN 136 145 POLY-PRO.  
 FT SEQUENCE 268 AA; 29568 MW; 6FA29C5798C75F91 CRC64;  
 Query Match 13.3%; Score 101.5; DB 1; Length 268;  
 Best Local Similarity 28.3%; Pred. No. 0.19;  
 Matches 41; Conservative 12; Mismatches 31; Indels 61; Gaps 9;  
 QY 4 ILQAPTAPS-----TIPGRGSGPEITFDPLPAPAPAPAPPSASGHRKESR 55  
 DB 132 VLSPPPPPSPPTPRSTPTIPHPRRSLPS-----PPSPSPSPSPSPSPSPS 178  
 QY 56 VLXPRVRLQVPEPNPAKLLFLLITVFCQILMAEGVPAIPLP-----PEDAPNA 108  
 DB 179 -----PRSTPI--PFRKR-----SPASPSPSLSKPSPSESP-- 211  
 QY 109 ASLAPTPVSPV--LEPNLTSEPSD 131  
 DB 212 -SLAPSPSDSVASLAP--SSSPSD 232  
 RESULT 11  
 ID WS14 MOUSE STANDARD; PRT; 864 AA.  
 AC Q99MZ3; Q99MZ9; Q99MZ0; Q99MZ1; Q99MZ2; Q99JMS;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Williams-Beuren syndrome chromosome region 14 protein homolog (Mlx  
 DE interactor).  
 GN WBSR14 OR MIO.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.; CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RX MEDLINE=21153101; PubMed=11230181;  
 RA Cairo S., Merla G., Urbinati F., Ballabio A., Raymond A.;  
 RT "WBSR14, a gene mapping to the Williams-Beuren syndrome deleted  
 RT region, is a new member of the Mlx transcription factor network.";  
 RL Hum. Mol. Genet. 10:617-627 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20241700; PubMed=10780788;  
 RA de Luis O., Valero M.C., Perez Jurado L.A.;  
 RT "WBSR14, a putative transcription factor gene deleted in Williams-  
 RT Beuren syndrome: complete characterisation of the human gene and the  
 RT mouse ortholog.";  
 RL Eur. J. Hum. Genet. 8:215-222 (2000).  
 CC -!- FUNCTION: Transcriptional repressor. Binds to the canonical and  
 CC non-canonical E box sequences 5'-CACGTG-3'.  
 CC -!- SUBUNIT: Binds DNA as a heterodimer with TCF4/MLX.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=5;  
 CC Name=1; Synonyms=Zeta;  
 CC -----

CC IsoId=Q99M23-1; Sequence=Displayed;  
CC Name=2; Synonyms=theta;  
CC IsoId=Q99M23-2; Sequence=VSP\_002174;  
CC Name=3; Synonyms=iota;  
CC IsoId=Q99M23-3; Sequence=VSP\_002177; VSP\_002178;  
CC Name=4; Synonyms=kappa;  
CC IsoId=Q99M23-4; Sequence=VSP\_002179; VSP\_002180;  
CC Name=5; Synonyms=eta;  
CC IsoId=Q99M23-5; Sequence=VSP\_002175; VSP\_002176;  
CC -!- TISSUE SPECIFICITY: Expressed in the ventricular and intermediate  
CC zones of the developing spinal cord of E12.5 embryos. In later  
CC embryos expressed in a variety of tissues.  
CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF245475; AAK20940.1; -;  
CC EMBL; AF245476; AAK20941.1; -;  
CC EMBL; AF245477; AAK20942.1; -;  
CC EMBL; AF245478; AAK20943.1; -;  
CC EMBL; AF245479; AAK20944.1; -;  
CC EMBL; AF156604; AAF68175.1; -;  
CC HSPF; P25912; IHG0.  
CC TRANSFAC; T05122; -;  
CC MGP; MGI:192799; Wbscr14.  
CC GO; GO:0005667; C:transcription factor complex; IDA.  
CC GO; GO:0016564; P:transcriptional repressor activity; IDA.  
CC GO; GO:0001022; P:negative regulation of transcription from P...; IDA.  
CC InterPro; IPR001092; HLH\_Basic.  
CC Pfam; PF001010; HLH; 1.  
CC SMART; SM00353; HLH; 1.  
CC PROSITE; P850888; HLH; 1.  
CC Transcription regulation; Repressor; Nuclear protein; DNA-binding;  
CC Alternative splicing.  
CC  
CC DRAIN 345 350 POLY-SER.  
CC FT DOMAIN 660 674 BASIC DOMAIN.  
CC FT DOMAIN 700 714 HELIX-LOOP-HELIX MOTIF.  
CC FT DOMAIN 715 736 LEUCINE-ZIPPER.  
CC FT VARSPLIC 58 79 Missing (in isoform 2).  
CC FT VARSPLIC 545 556 /FTID=VSP\_002174.  
CC FT VARSPLIC 545 556 ARPEALRPPTM -> VLVILVPSQA (in isoform  
CC 5).  
CC FT /FTID=VSP\_002175.  
CC FT VARSPLIC 557 864 Missing (in isoform 5).  
CC FT /FTID=VSP\_002176.  
CC FT VARSPLIC 699 744 YSKATLQKTAETILMLQGERAMQSEAAQRLDEIHELNA  
CC INICQ -> GLPTORTILVALAGQSNHASEDSGVHPDRAA  
CC GTGAGGAGAAAG (in isoform 3).  
CC FT /FTID=VSP\_002177.  
CC FT VARSPLIC 745 864 Missing (in isoform 3).  
CC FT /FTID=VSP\_002178.  
CC FT VARSPLIC 699 714 YSKATLQKTAETILM -> LPLANTENHIGARR (in  
CC isoform 4).  
CC FT /FTID=VSP\_002179.  
CC FT VARSPLIC 715 864 Missing (in isoform 4).  
CC FT /FTID=VSP\_002180.  
CC FT CONFLICT 67 67 D -> Y (IN REF. 2).  
CC FT CONFLICT 107 107 K -> N (IN REF. 2).  
CC FT CONFLICT 128 128 R -> I (IN REF. 2).  
CC FT CONFLICT 138 139 RK -> TR (IN REF. 2).  
CC FT CONFLICT 155 155 D -> H (IN REF. 2).  
CC FT CONFLICT 175 175 E -> D (IN REF. 2).  
CC FT CONFLICT 183 183 K -> V (IN REF. 2).  
CC FT CONFLICT 727 728 QQ -> HE (IN REF. 2).  
CC SEQUENCE 864 AA; 54874 MW; 736AFFE04C71B327 CRC64;  
SQ

Query Match

12.8%; Score 98; DB 1; Length 864;

Best Local Similarity 26.4%; Pred. No. 1.2;  
Matches 47; Conservative 13; Mismatches 46; Indels 72; Gaps 10;  
QY 10 PAPST-IPGRGSGPEITFDPLP-----EPA-----AAPAG-----RPSA-SRGHRKSS 53  
DB 455 PAPTFTVPTPGPGPGVPFSDHLEPHGLEVPFGPHFTVPGQPRCKPSGSPGQKAS 514  
QY 54 -----RVLYPRVVRQL--PYEEN----- 72  
DB 515 PPTLASATASPTATARDNNPCLQLLRAAKPQALEPTWPGTLRLPPSPQDTVSI 574  
QY 73 PAKRLFLILLITVFCQILMAEKGVPALPPEDAPNAASLAPTPVSPVLEPPNLTSPS 130  
DB 575 PRARAPF-----PPFAPTPPPPGPATLAP-PRSLVVPKARLSPPA 617  
RESULT 12  
WS14 HUMAN  
ID WS14 HUMAN STANDARD; PRT; 852 AA.  
AC Q9NP71; Q96B48; Q9BY03; Q9BY04; Q9BY05; Q9BY06; Q9Y2P3;  
DT 28-FEB-2003 (Rel. 41, created)  
DT 28-FEB-2003 (Rel. 41, last sequence update)  
DT 10-OCT-2003 (Rel. 42, last annotation update)  
DE Williams-Beuren syndrome chromosome region 14 protein (WS basic-helix-  
DE loop-helix leucine zipper protein) (WS-BHLH) (Mlx interactor).  
GN WBSCL14 OS MIO.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20241700; PubMed=10780798;  
RA de Luis O., Valero M.C., Perez Jurado L.A.;  
RT "WBSCL14, a putative transcription factor gene deleted in Williams-  
RT Beuren syndrome: complete characterisation of the human gene and the  
RT mouse ortholog.";  
PL Eur. J. Hum. Genet. 8:215-222(2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND SUBUNIT.  
RX PubMed=11230181;  
RA Cairo S., Merla G., Urbinati F., Ballabio A., Raymond A.;  
RT "WBSCL14, a gene mapping to the Williams-Beuren syndrome deleted  
RT region, is a new member of the Mlx transcription factor network.";  
PL Hum. Mol. Genet. 10:617-627(2001).  
RN [3]  
RP SEQUENCE OF 620-852 FROM N.A. (ISOFORM 4).  
RX MEDLINE=99075645; PubMed=9860302;  
RA Meng X., Lu X., Li Z., Green E.D., Massa H., Trask B.J., Morris C.A.,  
RA Keating M.T.;  
RT "Complete physical map of the common deletion region in Williams  
RT syndrome and identification and characterization of three novel  
RT genes.";  
PL Hum. Genet. 103:590-599(1998).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 6).  
RX TISSUE=EYE;  
RA MEDLINE=22389257; PubMed=12477932;  
RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustis T.B., Toshiyuki S., Carninci P., Frange C.,  
RA Rata S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,



RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: Transcriptional repressor. Binds to the canonical and  
 non-canonical E box sequences 5'-CACGTG-3' (By similarity).  
 CC -!- SUBUNIT: Binds DNA as a heterodimer with TCF4/MLX.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms-6;  
 CC Name=1; Synonyms=Alpha;  
 CC IsoId=Q9NP71-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Beta;  
 CC IsoId=Q9NP71-2; Sequence=VSP\_002171;  
 CC Name=3; Synonyms=Gamma;  
 CC IsoId=Q9NP71-3; Sequence=VSP\_002170;  
 CC Name=4; Synonyms=Delta;  
 CC IsoId=Q9NP71-4; Sequence=VSP\_002170, VSP\_002171;  
 CC Name=5; Synonyms=Epsilon;  
 CC IsoId=Q9NP71-5; Sequence=VSP\_002168, VSP\_002169;  
 CC Name=6;  
 CC IsoId=Q9NP71-6; Sequence=VSP\_002167, VSP\_002172, VSP\_002173;  
 CC -!- TISSUE SPECIFICITY: Expressed in liver, heart, kidney, cerebellum  
 and intestinal tissues.  
 CC -!- DISEASE: Rapininsufficiency of WSCR14 may be the cause of certain  
 cardiovascular and musculo-skeletal abnormalities observed in  
 Williams-Beuren syndrome (WBS) [MIM:194050], a rare developmental  
 disorder. It is a contiguous gene deletion syndrome involving  
 genes from chromosome band 7q11.23.  
 CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AF156673; AAF68176.1; -;  
 DR EMBL; AF156603; AAF68174.1; -;  
 DR EMBL; AF245470; AAK20935.1; -;  
 DR EMBL; AF245471; AAK20936.1; -;  
 DR EMBL; AF245472; AAK20937.1; -;  
 DR EMBL; AF245473; AAK20938.1; -;  
 DR EMBL; AF245474; AAK20939.1; -;  
 DR EMBL; AF056184; AAD28084.1; -;  
 DR EMBL; BC012925; AAH12925.1; -;  
 DR HSPF; P25912; IHLO.  
 DR TRANSFAC; T05121; -;  
 DR Genew; HGNC:12744; WSCR14.  
 DR MIM; 194050; -;  
 DR MIM; 605678; -;  
 DR GO; GO:0005667; C:transcription factor complex; NAS.  
 DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.  
 DR GO; GO:0006350; P:transcription; NAS.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR Pfam; PF00010; HLH; 1.  
 DR SMART; SM00353; HLH; 1.  
 DR PROSITE; PS00888; HLH; 1.  
 KW Transcription regulation; Repressor; Nuclear protein; DNA-binding;  
 KW Williams-Beuren syndrome; Alternative splicing.  
 FT DOMAIN 387 394 POLY-PRO.  
 FT DOMAIN 409 417 POLY-PRO.  
 FT DOMAIN 648 662 BASIC DOMAIN.  
 FT DOMAIN 688 702 HELIX-LOOP-HELIX MOTIF.  
 FT DOMAIN 703 724 LEUCINE-ZIPPER.  
 FT VARSPLIC 176 268 Missing (in isoform 6).  
 FT FTID=VSP\_002167.  
 FT QETVPEFCTFLPPTAP -> AVNGGQGGTSAPQALGL  
 (in isoform 5).  
 FT FTID=VSP\_002168.

FT VARSPLIC 576 852 Missing (in isoform 5).  
 FT FTID=VSP\_002169.  
 FT VARSPLIC 647 648 Missing (in isoform 3 and isoform 4).  
 FT FTID=VSP\_002170.  
 FT VARSPLIC 687 705 Missing (in isoform 2 and isoform 4).  
 FT FTID=VSP\_002171.  
 FT TV--ST (in isoform 6).  
 FT VARSPLIC 814 815 Missing (in isoform 6).  
 FT VARSPLIC 816 852 Missing (in isoform 6).  
 FT FTID=VSP\_002173.  
 FT CONFLICT 558 558 MISSING (IN REF. 4).  
 SQ SEQUENCE 852 AA; 93071 MW; D49E5C3D7C0A72EC CRC64;  
 Query Match 12.7%; Score 97; DB 1; Length 852;  
 Best Local Similarity 28.5%; Pred. NO. 1.4; Mismatches 52; Indels 50; Gaps 10;  
 Matches 45; Conservative 11;  
 QY 5 LQAPT---PAPSTIPGPRGSGPELFTPDLP---EPA-----AAPAGRPSA--SRGHR 50  
 Db 449 LPAPAAFTPTQSPSP---AFTPFPIELPLGYSEPAFGCFSGMRGKPPAPSPRG-Q 503  
 QY 51 KRERRVLYPRVVRQLEFVEEPNPAKRLFLLLTI-----VFCQILMASEG----- 95  
 Db 504 KASPTPLAPATASPTTAGSNFPC---LTQLITAAKPEQALEPPLVSVTLRSPSPQRT 560  
 QY 96 -----VPAIPPPEDAPNAASLAPFVSPVLEP 122  
 Db 561 VPEFCTFLPPTAPTPPPRPPGPATLAPS--RPLLPV 596  
 RESULT 13  
 VGF\_HUMAN STANDARD; PRT; 616 AA.  
 ID VGF\_HUMAN  
 AC 015240;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neurosecretory protein VGF precursor.  
 GN VGF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta;  
 RC MEDLINE=98008940; PubMed=9344675;  
 RX Canu N., Possenti R., Ricco A.S., Rocchi M., Levi A.;  
 RA "Cloning, structural organization analysis and chromosomal assignment  
 of the human gene for neurosecretory protein VGF.";  
 RL Genomics 45:443-446(1997)  
 CC -!- FUNCTION: May be involved in the regulation of cell-cell  
 interactions or in synaptogenesis during the maturation of the  
 nervous system (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Stored in secretory vesicles and then  
 secreted (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed in brain.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; Y12661; CAA73210.1; -;  
 CC Genew; HGNC:12684; VGF.  
 DR MIM; 602186; -;  
 DR GO; GO:0008083; F:growth factor activity; NAS.  
 DR Growth factor; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 616 NEUROSECRETORY PROTEIN VGF.

```

FT DOMAIN 353 447 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 615 AA; 67286 MW; CD1920610201BEB9 CRC64;

Query Match 12.6%; Score 96; DB 1; Length 616;
Best Local Similarity 24.1%; Pred. No. 1-2;
Matches 39; Conservative 13; Mismatches 40; Indels 70; Gaps 7;

QY 11 APTIPGRGSGPEI-----FTPDPLFPEAAAGRPFA 45
D5 11 :|||:||||:
44 AGDAVPGPKDGAPEVGRGARNSEPDGELFGVDPFRALAAVILQALDRPASPPA--PSG 101
QY 46 S-RGRHKSRLVPRVPRQ---LPVE-RPNPKELFLLLTVFCOILMAEGVAPL 100
D5 46 :|||:||||:
102 SQGPEEAEALLTVRSQTHSLPAGPEPA-----APPRPQ 141
QY 101 PPDAPNAASLAFTPVSPVLEPNTSEPSDYALDIATFLOQ 142
D5 101 :|||:||||:
142 TPENGPEA-----SDPSELEALASILQ 165

RESULT 14
MK07 MOUSE
ID MK07 MOUSE STANDARD; PRT; 806 AA.
AC Q9WV38;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mitogen-activated protein kinase 7 (EC 2.7.1.37) (Extracellular
DE signal-regulated kinase 5) (ERK-5) (BMK1 kinase).
GN MAPK7 OR ERK5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99403109; PubMed=10473620;
RA Kanakura S., Moriguchi T., Nishida E.;
RT "Activation of the protein kinase ERK5/BMK1 by receptor tyrosine
RT kinases. Identification and characterization of a signaling pathway
RT to the nucleus."
RL J. Biol. Chem. 274:26563-26571(1999).
CC -!- FUNCTION: MEK5 AND ERK5 INTERACT SPECIFICALLY WITH ONE ANOTHER AND
CC -!- NOT WITH MEK1/ERK1 OR MEK2/ERK2 PATHWAYS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.
CC -!- COFACTOR: Magnesium (By similarity).
CC -!- ENZYME REGULATION: Activated by tyrosine and threonine
CC phosphorylation (By similarity).
CC -!- DOMAIN: THE SECOND PROLINE-RICH REGION MAY INTERACT WITH ACTIN
CC TARGETING THE KINASE TO A SPECIFIC LOCATION IN THE CELL.
CC -!- FYM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN
CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
CC ROLE, IS ABSENT (BY SIMILARITY).
CC -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases. MAP
CC kinase subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB019373; BAA82039.1; -.
CC HSSP; P24941; LHCL.
CC DR MGI:1346347; Mapk7.
CC DR GO:0045765; P:regulation of angiogenesis; IMP.
CC DR InterPro: IPR003527; MAP kin.
CC DR InterPro: IPR000719; Prot kinase.
CC DR InterPro: IPR008271; Ser thr pkin AS.
CC DR InterPro: IPR002290; Ser thr_pkinase.
CC Pfam; PF00069; pkinase; 1.

```

```

DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TK; 1.
DR PROSITE; PS01351; MAPK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;
KW Phosphorylation.
FT DOMAIN 55 347 PROTEIN KINASE.
FT DOMAIN 55 347 PRO-RICH 1.
FT DOMAIN 434 465 POLY-ARG.
FT DOMAIN 521 524 POLY-RICH 2.
FT DOMAIN 576 700 PRO-RICH 2.
FT NP_BIND 61 69 ATP (BY SIMILARITY).
FT BINDING 84 84 ATP (BY SIMILARITY).
FT ACT_SITE 182 182 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 219 219 (BY SIMILARITY).
FT MOD_RES 221 221 PHOSPHORYLATION (ACTIVATES THE KINASE)
FT MOD_RES 221 221 (BY SIMILARITY).
FT MOD_RES 221 221 (BY SIMILARITY).
SQ SEQUENCE 806 AA; 87732 MW; E7CC41C4BBD0633 CRC64;

Query Match 12.5%; Score 95.5; DB 1; Length 806;
Best Local Similarity 29.1%; Pred. No. 1.8;
Matches 39; Conservative 12; Mismatches 66; Indels 17; Gaps 4;

QY 7 APTIPGRGSGPEI-----FTPDPLFPEAAAGRPFA 63
D5 7 :|||:||||:
580 APAPAPAPAPAPAPSA-----OPTSTGTGVSOSTGTLQAPAGSIPGPASQVCP 629
QY 64 RQLPYSEEN---PAKRLFLILLITVFCOILMAEGVAPLPDPEDAPNAASLAFTPVSPVL 120
D5 64 :|||:||||:
630 PPGFVPGAGIPAPLQTPAPSTLSASQSLVPSGLPGSGAPEVLVFPFS-GPPPPDPL 688
QY 121 EFNLTSEPSDYAL 134
D5 121 :|||:||||:
689 TPQPTSESPPVNL 702

RESULT 15
MKTA MOUSE
ID MKTA MOUSE STANDARD; PRT; 964 AA.
AC Q8KA76;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Myocardin-related transcription factor A (MYRF-A) (Megakaryoblastic
DE leukemia 1 protein homolog) (Basic SAP coiled-coil transcription
DE activator).
GN MKL1 OR BSAC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Spleen;
RX MEDLINE=22151085; PubMed=12019265;
RA Sasazuki T., Sawada Y., Sakon S., Kitamura T., Kishi T., Okazaki T.,
RA Katano M., Tanaka M., Watanabe M., Yagita H., Okumura K., Nakano H.;
RT "Identification of a novel transcriptional activator, BSAC, by a
RT functional cloning to inhibit tumor necrosis factor-induced cell
RT death."
RT J. Biol. Chem. 277:28853-28860(2002).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=C57BL/6;
RX MEDLINE=22317395; PubMed=12397177;
RA Wang D.-Z., Li S., Hockmeyer D., Sutherland L., Wang Z., Schratz G.,
RA Richardson J.A., Nordheim A., Olson E.N.;
RT "Potentiation of serum response factor activity by a family of
RT myocardin-related transcription factors."
RL Proc. Natl. Acad. Sci. U.S.A. 99:14855-14860(2002).
RN [3]

```

RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC STRAIN=C57BL/6J; TISSUE=Brain cortex;  
RX MEDLINE=22354683; PubMed=1246681;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Boro H., Kondo S.,  
RA Nikaudo I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanagiri A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Knadig D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Jendhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perte G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hiroane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs.";  
RL Nature 420:563-573 (2002).  
CC -!- FUNCTION: Transcriptional factor which uses the canonical single  
CC or multiple CARG boxes DNA sequence. Acts as a cofactor of serum  
CC response factor (SRF) with the potential to modulate SRF target  
CC genes. Suppresses TNF-induced cell death by inhibiting activation  
CC of caspases; its transcriptional activity is indispensable for the  
CC antapoptotic function. It may up-regulate antiapoptotic  
CC molecules, which in turn inhibit caspase activation.  
CC -!- SUBUNIT: Forms a ternary complex with SRF on DNA. Some authors  
CC (Ref.1) have found contradictory results, they could not  
CC demonstrated that it forms a complex with the CARG boxes, even in  
CC the presence of SRF.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Name=1;  
CC IsoId=Q8K4J6-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q8K4J6-2; Sequence=VSP\_007652;  
CC -!- TISSUE SPECIFICITY: Expressed in heart, brain, spleen, lung,  
CC liver, muscle, kidney and testis.  
CC -!- DEVELOPMENTAL STAGE: Detected throughout the embryo at 10.5 dpc;  
CC higher expression is found at 13.5 dpc in neural mesenchymal  
CC cells, skeletal muscle of the tongue, and epithelial cells of the  
CC colon and small intestine; at 15.5 dpc, expression in epithelial  
CC cells of lung, kidney, bladder, and colon is also detected.  
CC -!- DOMAIN: The N-terminal region is required for nuclear localization  
CC and the C-terminal region mediates transcriptional activity.  
CC -!- SIMILARITY: Contains 3 RPEL repeats.  
CC -!- SIMILARITY: Contains 1 SAP domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF385592; AAM94259.1; -.  
CC EMBL; AF532597; AAN33041.1; -.  
CC EMBL; AK044188; BAC31809.1; -.

DR EMBL; AK089416; BAC40873.1; -.  
DR MGD; MGI:238495; Mkl1.  
DR GO; GO:0005634; C:nucleus; IDA.  
DR GO; GO:0008189; F:apoptosis inhibitor activity; IDA.  
DR GO; GO:0016563; P:transcriptional activator activity; IDA.  
DR GO; GO:0006350; P:transcription; IDA.  
DR InterPro; IPR004018; RPEL\_repeat.  
DR InterPro; IPR003034; SAP.  
DR Pfam; PF02755; RPEL; 3.  
DR Pfam; PF02037; SAP; 1.  
DR SMART; SM00707; RPEL; 3.  
DR SMART; SM00513; SAP; 1.  
DR PROSITE; PS50800; SAP; 1.  
KW Transcription regulation; Nuclear protein; Coiled coil; Repeat;  
KW Alternative splicing.  
FT REPEAT 15 40  
FT REPEAT 59 84  
FT REPEAT 103 128  
FT DOMAIN 299 325  
FT DOMAIN 385 419  
FT DOMAIN 552 600  
FT DOMAIN 712 844  
FT VARSPIC 1 35  
FT CONFLICT 53 53  
FT CONFLICT 724 724  
FT CONFLICT 728 728  
FT SEQUENCE 964 AA; 102545 MW; AFAEA328A1860CE5 CRC64;  
SQ  
Query Match 12.5%; Score 95.5; DB 1; Length 964;  
Best Local Similarity 24.7%; Pred. No. 2.1;  
Matches 43; Conservative 24; Mismatches 78; Indels 29; Gaps 5;  
QY 1 TMTLQATPAPSTIPQPRGSGPIFTFDLPDPAAPAGPSPASRGHKKSRVLYER 60  
Db 208 TLTFLAQPELPAPLPPSLANGSIVPTAKPAPTLIKQSPKASXKSQSKAKELPK 267  
QY 61 VVR-----RQLPVEEPNPAKRL-----LFLLLTVFQILMAEAG----- 95  
Db 268 VKLKYHQVIPPQDKQKAPAMDSVAKILQQQQLFLOQLINQOQQOQQOQHYNQAI 327  
QY 96 VPAPLPDPDAPNAASLAPTP---VSPVLEPNLTSPSYALDLSTFLQQHPAA 146  
Db 328 LPAP-PKPSAETPGSSAPTFPSRLSTSSSPSGTPTGPGSLAQSGSTAAAKPGA 380  
Search completed: February 25, 2004, 05:27:11  
Job time : 19 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:24:25 ; Search time 39 Seconds  
(without alignments)  
1189.261 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_10\_156

Perfect score: 764

Sequence: 1 TMTILQAPTPAPSTIPGRR.....EPDVALDLSTFIQQHPAPF 147

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_plage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description         |
|------------|-------|-------------|--------|-----------|---------------------|
| 1          | 757   | 99.1        | 156    | 6 Q7YR42  | Q7YR42 pan troglod  |
| 2          | 712.5 | 99.3        | 193    | 4 Q75353  | Q75353 homo sapien  |
| 3          | 540   | 70.7        | 160    | 11 Q91VZ5 | Q91VZ5 mus musculus |
| 4          | 124.5 | 16.3        | 708    | 16 Q7WAY9 | Q7WAY9 bordetella   |
| 5          | 122.5 | 16.0        | 423    | 4 Q9NXH8  | Q9NXH8 homo sapien  |
| 6          | 120.5 | 15.8        | 696    | 16 Q7VY14 | Q7VY14 bordetella   |
| 7          | 115.5 | 15.1        | 743    | 16 Q7WNF3 | Q7WNF3 bordetella   |
| 8          | 113   | 14.8        | 1159   | 16 Q7U7U9 | Q7U7U9 synecococc   |
| 9          | 112.5 | 14.7        | 649    | 5 Q9V884  | Q9V884 dysophila    |
| 10         | 111.5 | 14.6        | 554    | 6 Q95L89  | Q95L89 bos taurus   |
| 11         | 111.5 | 14.6        | 580    | 6 Q8WML4  | Q8WML4 bos taurus   |
| 12         | 109.5 | 14.3        | 668    | 5 Q8MXU8  | Q8MXU8 caenorhabdi  |
| 13         | 109.5 | 14.3        | 3247   | 12 Q65553 | Q65553 bovine herp  |
| 14         | 109   | 14.3        | 3325   | 12 Q91BT9 | Q91BT9 turkey herp  |
| 15         | 108.5 | 14.2        | 233    | 2 Q9L568  | Q9L568 streptococ   |
| 16         | 108.5 | 14.2        | 3179   | 12 Q8V2A4 | Q8V2A4 human herpe  |

|    |       |      |      |           |                    |
|----|-------|------|------|-----------|--------------------|
| 17 | 108   | 14.1 | 224  | 2 Q8GNS8  | Q8GNS8 streptococ  |
| 18 | 108   | 14.1 | 380  | 11 Q921D7 | Q921D7 mus musculu |
| 19 | 108   | 14.1 | 426  | 11 Q8BH02 | Q8BH02 mus musculu |
| 20 | 107.5 | 14.1 | 160  | 6 Q8SQ36  | Q8SQ36 bos taurus  |
| 21 | 107.5 | 14.1 | 203  | 16 Q9RCX9 | Q9RCX9 streptomyc  |
| 22 | 107.5 | 14.1 | 542  | 12 Q91CH9 | Q91CH9 macropodid  |
| 23 | 107.5 | 14.1 | 502  | 10 Q8LGS4 | Q8LGS4 arabidopsi  |
| 24 | 107   | 14.0 | 426  | 2 Q9LAY5  | Q9LAY5 streptococ  |
| 25 | 107   | 14.0 | 496  | 12 Q98457 | Q98457 paramecium  |
| 26 | 107   | 14.0 | 875  | 2 Q9F2B0  | Q9F2B0 thibacilli  |
| 27 | 106.5 | 13.9 | 160  | 6 Q8SQ37  | Q8SQ37 bos taurus  |
| 28 | 106.5 | 13.9 | 581  | 5 Q20517  | Q20517 caenorhabdi |
| 29 | 106.5 | 13.9 | 1394 | 4 Q72744  | Q72744 homo sapien |
| 30 | 106.5 | 13.9 | 1668 | 4 Q15026  | Q15026 homo sapien |
| 31 | 106.5 | 13.9 | 2971 | 4 Q9V5L9  | Q9V5L9 homo sapien |
| 32 | 106   | 13.9 | 130  | 10 Q9FKQ0 | Q9FKQ0 arabidopsi  |
| 33 | 106   | 13.9 | 1029 | 12 Q9YMX0 | Q9YMX0 lymantria d |
| 34 | 106   | 13.9 | 1709 | 4 Q15047  | Q15047 homo sapien |
| 35 | 105.5 | 13.8 | 249  | 2 Q9L570  | Q9L570 streptococ  |
| 36 | 105   | 13.7 | 309  | 12 Q8UZB4 | Q8UZB4 grapevine f |
| 37 | 105   | 13.7 | 549  | 12 Q89370 | Q89370 paramecium  |
| 38 | 105   | 13.7 | 3342 | 12 Q956N3 | Q956N3 turkey herp |
| 39 | 104.5 | 13.7 | 243  | 2 Q9L567  | Q9L567 streptococ  |
| 40 | 104.5 | 13.7 | 244  | 2 Q9L565  | Q9L565 streptococ  |
| 41 | 104.5 | 13.7 | 552  | 10 Q96343 | Q96343 brassica na |
| 42 | 104.5 | 13.7 | 616  | 12 Q96716 | Q96716 chlorella v |
| 43 | 104.5 | 13.7 | 2066 | 12 P89920 | P89920 oat blue dw |
| 44 | 104   | 13.6 | 351  | 10 Q39492 | Q39492 chlamydomon |
| 45 | 104   | 13.6 | 395  | 2 Q9LAY2  | Q9LAY2 streptococ  |

## ALIGNMENTS

RESULT 1

Q7YR42 PRELIMINARY; PRT; 156 AA.  
AC Q7YR42;  
DT 01-OCT-2003 (TREMELrel. 25, Created)  
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE IEX-1 protein.  
GN IEX-1.  
OS Pan troglodytes (Chimpanzee).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.  
OX NCBI\_TaxID=9598;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22709134; PubMed=12799463;  
RA Anzai T., Shiina T., Kimura N., Yanagiya K., Kohara S., Shigenari A., Yamagata T., Kulski J.K., Naruse T.K., Fujimori Y., Fukuzumi Y., Yamazaki M., Tashiro H., Iwamoto C., Umebara Y., Imanishi T., Meyer A., Ikeo K., Gojochi T., Bahram S., Inoko H.;  
RA "Comparative sequencing of human and chimpanzee MHC class I regions  
RT unveils insertions/deletions as the major path to genomic  
RT divergence".  
RT Proc. Natl. Acad. Sci. U.S.A. 100:7708-7713(2003).  
RL ENBL; AB100084; BAC78173.1; -.  
SQ SEQUENCE 156 AA; 16929 MW; 83C0670DCAC09650 CRC64;

Query Match 99.1%; Score 757; DB 6; Length 156;  
Best Local Similarity 99.3%; Pred. No. 3.9e-62;  
Matches 146; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

|    |    |  |     |
|----|----|--|-----|
| Qy | 1  | TMTILQAPTPAPSTIPGRRSGFEITFDPLPFAAAPAGRPSASGRHKRSRVLYPR   | 60  |
| Ds | 10 | TMTILQAPTPASTNPGRSGFEITFDPLPFAAAPAGRPSASGRHKRSRVLYPR     | 69  |
| Qy | 61 | VRRQLPVSEPNPAKRLIFLLITIVFCQLMAEGVPAFLPPEDAPNAASLAPTPSPVL | 120 |
| Ds | 70 | VRRQLPVSEPNPAKRLIFLLITIVFCQLMAEGVPAFLPPEDAPNAASLAPTPSPVL | 129 |

```

QY 121 EPPNLTSPSYDLDLSTFLOQHPAAF 147
DB 130 EPPNLTSPSYDLDLSTFLOQHPAAF 156

RESULT 2
O75353 PRELIMINARY; PRT; 193 AA.
ID O75353
AC O75353;
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE ANTI-death protein.
GN IEX-1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98369175; PubMed=9703517;
RA Wu M.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell
RT survival.";
RL Science 281:998-1001(1998).
DR EMBL; AF039067; AAC32558.1; -.
DR EMBL; AF071596; AAC72344.1; -.
DR GO; GO:0008189; P:apoptosis inhibitor activity; NAS.
DR GO; GO:0006916; P:anti-apoptosis; NAS.
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3FFBC7C57 CRC64;

Query Match 93.3%; Score 712.5; DB 4; Length 193;
Best Local Similarity 77.7%; Pred. NO. 6e-58;
Matches 143; Conservative 1; Mismatches 3; Indels 37; Gaps 1;

QY 1 TWTLQAPTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPR 60
DB 10 TWTLQAPTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPR 59
QY 61 TWTLQAPTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPR 83
DB 70 SGEYRSGHSCALPWSHWTTLQSLDTCLAFCLPLVYRQLVPEPNPAKLLFLIT 129
QY 84 IVFCQILMAEGVPAPLPDPDAPNAASLAPTSPVLEPNLTSPSYDLDLSTFLOQ 143
DB 130 IVFCQILMAEGVPAPLPDPDAPNAASLAPTSPVLEPNLTSPSYDLDLSTFLOQ 189
QY 144 PPAF 147
DB 190 PPAF 193

RESULT 3
Q91VZ5 PRELIMINARY; PRT; 160 AA.
ID Q91VZ5
AC Q91VZ5;
DT 01-DEC-2001 (TREMELrel. 19, Created)
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
DE Similar to immediate early response 3.
GN IER3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Breast tumor;
RA Strausberg R.;
RN Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Body;

```

```

RX MEDLINE=22354583; PubMed=12466851;
RA The FANTOM Consortium,
RA "The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; BC006950; AAH06950.1; -.
DR EMBL; AK051003; BAC34493.1; -.
DR MGD; MGI:104814; Ier3.
SQ SEQUENCE 160 AA; 17655 MW; 0666DF96E751FCF4 CRC64;

Query Match 70.7%; Score 540; DB 11; Length 160;
Best Local Similarity 72.5%; Pred. NO. 3.7e-42;
Matches 111; Conservative 8; Mismatches 26; Indels 8; Gaps 2;

QY 1 TWTLQAPTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPR 60
DB 10 TWTLQAPTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPR 69
QY 61 VVERQLPVEEPNPAKLLFLITIVFCQILMAEGVPAPLPDPDAPNAASLAPTSPV 117
DB 70 VVERQLPVEEPNPAKLLFLITIVFCQILMAEGVPAPLPDPDAPNAASLAPTSPV 127
QY 118 ---PVLEPNLTSPSYDLDLSTFLOQHPAAF 147
DB 128 TAPVLEPNLTSPSYDLDLSTFLOQHPAAF 160

RESULT 4
Q7WAY9 PRELIMINARY; PRT; 708 AA.
ID Q7WAY9
AC Q7WAY9;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE DNA polymerase III subunit Tau (EC 2.7.7.7).
GN DNAX OR DNAX OR DNAX OR BPP1221.
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebatilla M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.B., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moulé S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640426; CAE36523.1; -.
KW Nucleotidyltransferase; Transferase; Complete proteome.
SQ SEQUENCE 708 AA; 74513 MW; C0ACC464C151BCE9 CRC64;

Query Match 16.3%; Score 124.5; DB 16; Length 708;
Best Local Similarity 32.8%; Pred. NO. 0.003;
Matches 38; Conservative 6; Mismatches 43; Indels 29; Gaps 5;

QY 8 PTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPRVVRQLP 67
DB 416 PTPAPSTIPGPRGSGPIFTFDLPPEAAAPAGPSASGHKRSRRLVYPRVVRQLP 459
QY 68 VVERNPAKLLFLITIVFCQILMAEGVPAPLPDPDAPNAASLAPTSPVLEP 122
DB 460 VAEPEP-----VAAFTQARAAPAFVFPVWEDLEAVA-APAPQAPAAAP 503

```



```

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN SYN0881.
OS Synchococcus sp. (strain WH102).
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.
OX NCBI_TaxID=84588;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22925697; PubMed=12917641;
RA Palenik B., Brahamsha B., Larimer F.W., Land M., Kausner L., Chain P.,
RA Lamerdin J., Regala W., Allen E.E., McCarren J., Paulsen I.,
RA Dufresne A., Partensky F., Webb E.A., Waterbury J.;
RT "The genome of a motile marine Synchococcus."
RL Nature 424:1037-1042(2003).
DR EMBL; BX563691; CAB07396.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 1159 AA; 127276 MW; 7448D814D090C3F6 CRC64;

Query Match 14.8%; Score 113; DB 16; Length 1159;
Best Local Similarity 29.4%; Pred. No. 0.058;
Matches 35; Conservative 7; Mismatches 53; Indels 24; Gaps 4;

QY 8 PTPAPSTIPGPRGSGFEITFDPLPEPAAPAGPSASRGHRKSRVLYRVVRQLP 67
DB 585 PTPAPSTIPGPRGSGFEITFDPLPEPAAPAGPSASRGHRKSRVLYRVVRQLP 641
QY 68 VEEPNAKRLFLILATVFCQILMAEGVPALPPEDAPNAASLAPTPSVLEPFNLT 126
DB 642 TREPTAP-----ATTPTPEIPVPTPEQDS-----SPTPTG---NPNTT 679

RESULT 9
QV884 PRELIMINARY; PRT; 649 AA.
AC QV884
AC QV884;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE CG18635 protein.
GN CG18635
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20186006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., McKus G.L.G.,
RA Abriel J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA De Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Durkin B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischman W.,
RA Rosier C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush P., Karpen G.H., Ke Z., Kennison C.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

```

```

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy D., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskein D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Wodtke T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003802; AA057791.1; -.
DR Flybase; FBgn0034279; CG18635.
SQ SEQUENCE 649 AA; 74414 MW; 6471AC52B49F9AE9 CRC64;

Query Match 14.7%; Score 112.5; DB 5; Length 649;
Best Local Similarity 25.8%; Pred. No. 0.035;
Matches 46; Conservative 26; Mismatches 67; Indels 39; Gaps 7;

QY 1 TWTILQA--PTTAPSTIPGPRGSGFEITFDPLPEPAAPAGPSASRGHRKSRVLY 58
DB 342 TIVLDSITETPAP---PPPTAPPELPVLKPTEDDPTPLRLTSLDNEHKSRVTR 398
QY 59 P-----RVVRQLPVE--PNPAKRLFLLL-----TIVFCQILMAEGVPAP 99
DB 399 PLQLETFKMLLVNAQLVKEIVNPPKLLIGRLKDEDEPHT-----VLPQSPKPS 453
QY 100 LPPEAPNAASLAPTPSVLEPFNLTSEPSDY-----ALDLSTFLQHPAAF 147
DB 454 LTPKQVVDYFPFRETQIVNGIQDDFAGRTVSPFGWIAFTWRLSLSTFCVYPKAF 511

RESULT 10
QV5189 PRELIMINARY; PRT; 554 AA.
AC QV5189
AC QV5189;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Mucin (Fragment).
GN MUC3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Sacchi P., Rasser R., Rosati S.;
RT "Molecular analysis of MUC1 polymorphism in cattle."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF39757; AL28023.1; -.
DR InterPro; IPR002965; P-rich extensin.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PR01217; PRICHEXTENSIN.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS00024; SEA; 1.
DR NON_TER 554
SQ SEQUENCE 554 AA; 55508 MW; B441CD140093AFFD CRC64;

Query Match 14.6%; Score 111.5; DB 6; Length 554;
Best Local Similarity 28.3%; Pred. No. 0.037;
Matches 41; Conservative 10; Mismatches 41; Indels 53; Gaps 7;

QY 1 TWTILQA--PTTAPSTIPGPRGSGFEITFDPLPEPA-----APAGPSAS 46
DB 61 TWTILASSPAPSPAASPGHDGASTP---TSSPAPSPAASPGHDGASTPSPAAS 117

```

```
QY 47 RGHKRSRVLYPRVVRQLPVEPNPAKELLFLLLTIVFCQILMAEKGVPAPLPDPAD 106
Db 118 PGHDCAS-----TPTSSPAPS-----PAASPGHDGA 143
QY 107 NAASLAPTPVSPVLEP-FNLTSEPS 130
Db 144 STPTSSPAP-SPAASPGHNGTSPT 167

RESULT 11
Q8MWL4
ID Q8MWL4 PRELIMINARY; PRT; 580 AA.
AC Q8MWL4
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE MUC1 protein precursor.
GN MUC1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Holstein Dairy cow; TIGSUS=Lactating mammary gland;
RX MEDLINE=21672493; PubMed=11814015;
RA Pallesen L.T., Andersen M.H., Nielsen R.L., Berglund L.,
RA Rasmussen L.K., Petersen T.E., Rasmussen J.T.;
RT "Purification of MUC1 from bovine milk-fat globules and
RT characterization of a corresponding full-length cDNA Clone.";
RL J. Dairy Sci. 84:2591-2598(2001).
DR EMEL; AJ400824; CAC61810.1; -.
DR InterPro; IPR002965; P-rich extensn.
DR InterPro; IPR000082; SEA_domain.
DR Pfam; PF01390; SEA; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00200; SEA; 1.
DR PROSITE; PS50024; SEA; 1.
KW Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 580 AA; 58091 MW; E91C13984AF7D757 CRC64;

Query Match 14.6%; Score 111.5; DB 6; Length 580;
Best Local Similarity 28.3%; Pred. No. 0.039;
Matches 41; Conservative 10; Mismatches 41; Indels 53; Gaps 7;

QY 1 TWTLIQA--PTPAPSTIPGRGSGPEIFTFPLPEPAA-----APAGRPAS 46
Db 61 TWTLTASSAPSPASPGHDGATP---TSSPAPSPASPGHDGASTTSSPAPSPAS 117
QY 47 RGHKRSRVLYPRVVRQLPVEPNPAKELLFLLLTIVFCQILMAEKGVPAPLPDPAD 106
Db 118 PGHDCAS-----TPTSSPAPS-----PAASPGHDGA 143
QY 107 NAASLAPTPVSPVLEP-FNLTSEPS 130
Db 144 STPTSSPAP-SPAASPGHNGTSPT 167

RESULT 12
Q8MXU8
ID Q8MXU8 PRELIMINARY; PRT; 668 AA.
AC Q8MXU8;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein K08D12.6.
GN K08D12.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.

QY 47 RGHKRSRVLYPRVVRQLPVEPNPAKELLFLLLTIVFCQILMAEKGVPAPLPDPAD 106
Db 118 PGHDCAS-----TPTSSPAPS-----PAASPGHDGA 143
QY 107 NAASLAPTPVSPVLEP-FNLTSEPS 130
Db 144 STPTSSPAP-SPAASPGHNGTSPT 167

RESULT 13
Q65553
ID Q65553 PRELIMINARY; PRT; 3247 AA.
AC Q65553;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UL36.
GN UL36.
OS Bovine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M., Vleck C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
RA Thiry E., Paces V.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA Laboussiere S., Misra V., Vleck C., Paces V.;
RL Vet. Microbiol. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=JURA;
RX MEDLINE=97164286; PubMed=9010999;
RA Schwytzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
```

```
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Waterston R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wu X., Gattung S.;
RT "The sequence of C. elegans cosmid K08D12.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMEL; AC008672; AAM98004.1; -.
DR WormPep; K08D12.6; CE21047.
KW Hypothetical protein.
SQ SEQUENCE 668 AA; 62778 MW; FB9ACC5739DC398A CRC64;

Query Match 14.3%; Score 109.5; DB 5; Length 668;
Best Local Similarity 31.2%; Pred. No. 0.069;
Matches 35; Conservative 6; Mismatches 50; Indels 21; Gaps 3;

QY 8 PTPAPSTIPGRGSGPEIFTFPLPEPAAAPAGRPASRGHRSRVLYPRVVRQLP 67
Db 118 PAPAAPAPDCGSA-----APAAAPAAAPATDSGVSKENAYGDEQUTPAP 168
QY 68 VEEPNAKRLFLLLTIVFCQILMAEKGVPAPLP-PEDAPNAAIAPTFVSP 118
Db 169 AAAPAPAD-----APVEQAPVAVPAPAPAAAPDCGSAAPAPAP 209

RESULT 13
Q65553
ID Q65553 PRELIMINARY; PRT; 3247 AA.
AC Q65553;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE UL36.
GN UL36.
OS Bovine herpesvirus 1.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10320;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M., Vleck C., Lowery D.E., Bello L.J., Meyer G., Misra V.,
RA Thiry E., Paces V.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
RA Laboussiere S., Misra V., Vleck C., Paces V.;
RL Vet. Microbiol. 0:0-0(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=COOPER;
RA Schwytzer M.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-179 FROM N.A.
RC STRAIN=JURA;
RX MEDLINE=97164286; PubMed=9010999;
RA Schwytzer M., Styger D., Vogt B., Lowery D.E., Simard C.,
```



RA LaBoissiere S., Misra V., Vleck C., Paces V.;  
 RT "Gene contents in a 31-kb segment at the left genome end of bovine  
 RL herpesvirus-1";  
 RN Vet. Microbiol. 53:67-77(1996).  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JURA;  
 RA Schwytzer M., Vleck C., Lowery D.E., Bello L.J., Meyer G., Misra V.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JURA;  
 RA Schwytzer M.;  
 RL Submitted (SBB-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; 278205; CAB01605.1; -;  
 DR EMBL; AJ004801; CA06097.1; -;  
 DR InterPro; IPR006928; Herpes\_teg\_N.  
 DR InterPro; IPR005210; Herpes\_UL36.  
 DR Pfam; PF04843; Herpes\_teg\_N; 1.  
 DR Pfam; PF03586; Herpes\_UL36; 1.  
 SQ SEQUENCE 3247 AA; 332190 MW; 3AEAA72F8F001F6A CRC64;  
  
 Query Match 14.3%; Score 109.5; DB 12; Length 3247;  
 Best Local Similarity 31.4%; Pred. No. 0.35; Indels 49; Gaps 6;  
 Matches 38; Conservative 6; Mismatches 28;  
  
 QY 8 PTPAPSTIPGRRGSGPFIETFDLPPEPAAAGPAGPSASRGHKKRRVLYRVVRRQLP 67  
 DB 2663 PTPAPSTAPVA-----PPLPPALTPA-----P 2690  
  
 QY 68 VVEENPAKRI-LPFLITVFCQILMAEGVPALP-----PEDAPNAASLAPTFVSPVLE 121  
 DB 2691 TEATPAPPLPAPITVL-----VPAEPAPAPIPAPAPITAP-APTAPPLPP 2739  
  
 QY 122 P 122  
 DB 2740 P 2740  
  
 RESULT 14  
 QYBT9 PRELIMINARY; PRT; 3325 AA.  
 AC QYBT9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Major tegument protein.  
 GN UL36.  
 OS Turkey herpesvirus.  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Marek's disease-like viruses.  
 OX NCBI\_TaxID=10390;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GA;  
 RX MEDLINE=2237304; PubMed=1315048;  
 RA Jones D., Lee L., Liu J.L., Kung H.J., Tillotson J.K.;  
 RT "Marek disease virus encodes a basic-leucine zipper gene resembling  
 RT the fos/jun oncogenes that is highly expressed in lymphoblastoid  
 RT tumors";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:6091-6096(2000).  
 DR EMBL; AF147806; AAF66771.1; -;  
 DR InterPro; IPR006928; Herpes\_teg\_N.  
 DR InterPro; IPR005210; Herpes\_UL36.  
 DR Pfam; PF04843; Herpes\_teg\_N; 1.

DR Pfam; PF03586; Herpes\_UL36; 1.  
 SQ SEQUENCE 3325 AA; 365695 MW; 5D4A84EF719BB9FD CRC64;  
  
 Query Match 14.3%; Score 109; DB 12; Length 3325;  
 Best Local Similarity 25.2%; Pred. No. 0.4;  
 Matches 31; Conservative 14; Mismatches 52; Indels 26; Gaps 3;  
  
 QY 7 APTAPSTIPGRRGSGPFIETFDLPPEPAAAGPAGPSASRGHKKRRVLYRVVRRQL 66  
 DB 2802 SPAPKSPAPKPPDPD---FKPSAPKSPASKPSKSPASKPPADPKPS 2858  
  
 QY 67 PVVEENPAKLLPFLITVFCQILMAEGVPALPPEPAPNAASLAPTF-----VSPV 119  
 DB 2859 PAPKPPPP-----TPDSKSPAPKSPASKPILPVFPFNSDKSPV 2902  
  
 QY 120 LEP 122  
 DB 2903 PNF 2905  
  
 RESULT 15  
 QYLT568 PRELIMINARY; PRT; 233 AA.  
 AC QYLT568;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE PSPA (Fragment).  
 GN PSPA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=39;  
 RX MEDLINE=20472698; PubMed=11015380;  
 RA Beall B., Cherardi G., Facklam R.R., Hollingshead S.K.;  
 RT "Pneumococcal pspA Sequence Types of Prevalent Multiresistant  
 RT Pneumococcal Strains in the United States and of Internationally  
 RT Disseminated Clones";  
 RL J. Clin. Microbiol. 38:3663-3669(2000).  
 DR EMBL; AF255902; AAF70092.1; -;  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR InterPro; IPR003345; M\_repeat.  
 DR Pfam; PF02370; M; 3. 1  
 FT NON\_TER 1 233  
 FT NON\_TER 233 233  
 SQ SEQUENCE 233 AA; 24514 MW; D5C494019C45BFE2 CRC64;  
  
 Query Match 14.2%; Score 108.5; DB 2; Length 233;  
 Best Local Similarity 28.1%; Pred. No. 0.029;  
 Matches 32; Conservative 10; Mismatches 35; Indels 37; Gaps 5;  
  
 QY 6 QAPTAPSTIPGRRGSGPFIETFDLPPEPAAAGPAGPSASRGHKKRRVLYRVVRRQ 65  
 DB 146 EAPAPKAPAPKAPAPK-----PAPAPKAPAPKAPAP-----KEAPAPKP 190  
  
 QY 66 LPVEENPAKLLPFLITVFCQILMAEGVPALP-PEDAPNAASLAPTFVSP 118  
 DB 191 APAPKAPAPK-----PAPAPKAPAPKAPAP-APKPTP 223  
  
 Search completed: February 25, 2004, 05:28:04  
 Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:29:16 ; Search time 54 Seconds  
(without alignments)  
449.983 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442

Sequence: 1 VRQLPVFPNPKLLFL.....RPSDYALDLSTFLQHPAFA 86

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq 29Jan04:\*

- 1: Geneseqp1980s:\*
- 2: Geneseqp1990s:\*
- 3: Geneseqp2000s:\*
- 4: Geneseqp2001s:\*
- 5: Geneseqp2002s:\*
- 6: Geneseqp2003as:\*
- 7: Geneseqp2003bs:\*
- 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID      | Description        |
|------------|-------|-------------|--------|------------|--------------------|
| 1          | 442   | 100.0       | 156    | 2 AAW36006 | Raw36006 Human Fch |
| 2          | 442   | 100.0       | 156    | 3 AAV45017 | Aay45017 Protein e |
| 3          | 442   | 100.0       | 156    | 5 AAU78232 | Aau78232 Human PRO |
| 4          | 442   | 100.0       | 156    | 7 ADD46473 | Add46473 Human PRO |
| 5          | 442   | 100.0       | 156    | 7 ADD44772 | Add44772 Human PRO |
| 6          | 442   | 100.0       | 156    | 7 ADD44775 | Add44775 Human PRO |
| 7          | 442   | 100.0       | 156    | 7 ADD44769 | Add44769 Human PRO |
| 8          | 442   | 100.0       | 156    | 7 ADD44778 | Add44778 Human PRO |
| 9          | 442   | 100.0       | 166    | 4 AAG74873 | Aag74873 Human col |
| 10         | 423   | 95.7        | 82     | 4 AAE03933 | Aae03933 Human gen |
| 11         | 421   | 95.2        | 83     | 4 AAE03976 | Aae03976 Human gen |
| 12         | 259   | 58.6        | 83     | 4 AAG75952 | Aag75952 Human col |
| 13         | 84.5  | 19.1        | 923    | 4 AAE11773 | Aae11773 Human kin |
| 14         | 84.5  | 19.1        | 926    | 4 AAB5631  | Aab5631 Novel pro  |
| 15         | 84.5  | 19.1        | 950    | 7 ADD14168 | Add14168 Human src |
| 16         | 84.5  | 19.1        | 958    | 4 ABUS2390 | Abu2390 Human sig  |
| 17         | 79.5  | 18.0        | 1596   | 3 AAY68821 | Aay68821 Amino aci |
| 18         | 77    | 17.4        | 835    | 5 ABP43616 | Abp43616 19 cosmid |
| 19         | 77    | 17.4        | 2424   | 4 ABB58924 | Abb58924 Drosophil |
| 20         | 76    | 17.2        | 801    | 4 ABB58990 | Abb58990 Drosophil |
| 21         | 75.5  | 17.1        | 242    | 4 AAM16455 | Aam16455 Peptide # |
| 22         | 75.5  | 17.1        | 242    | 4 ABE35443 | Abb35443 Peptide # |
| 23         | 75.5  | 17.1        | 242    | 4 AAM28950 | Aam28950 Peptide # |
| 24         | 75.5  | 17.1        | 242    | 4 ABB30277 | Abb30277 Peptide # |
| 25         | 75.5  | 17.1        | 242    | 4 ABB20882 | Abb20882 Protein # |

|    |      |      |      |            |                    |
|----|------|------|------|------------|--------------------|
| 26 | 75.5 | 17.1 | 242  | 4 AAM68643 | Aam68643 Human bon |
| 27 | 75.5 | 17.1 | 242  | 4 AAM58266 | Aam58266 Human bra |
| 28 | 75.5 | 17.1 | 242  | 4 AAG50303 | Agg50303 Human liv |
| 29 | 75.5 | 17.1 | 242  | 4 AAM04184 | Aam04184 Peptide # |
| 30 | 75.5 | 17.1 | 242  | 5 ABB38222 | Abb38222 Human pep |
| 31 | 74   | 16.7 | 955  | 7 ADE59235 | Ades9235 Rat Prote |
| 32 | 74   | 16.7 | 1040 | 4 ABB65945 | Abb65945 Drosophil |
| 33 | 72.5 | 16.4 | 111  | 4 AAM19703 | Aam19703 Peptide # |
| 34 | 72.5 | 16.4 | 111  | 4 ABB39430 | Abb39430 Peptide # |
| 35 | 72.5 | 16.4 | 111  | 4 AAM32953 | Aam32953 Peptide # |
| 36 | 72.5 | 16.4 | 111  | 4 ABB24202 | Abb24202 Protein # |
| 37 | 72.5 | 16.4 | 111  | 4 AAM72722 | Aam72722 Human bon |
| 38 | 72.5 | 16.4 | 111  | 4 AAM60110 | Aam60110 Human bra |
| 39 | 72.5 | 16.4 | 111  | 4 AAG54420 | Agg54420 Human liv |
| 40 | 72.5 | 16.4 | 111  | 5 ABB42546 | Abb42546 Human pep |
| 41 | 72.5 | 16.4 | 420  | 4 ABB26382 | Abb26382 Novel hum |
| 42 | 72   | 16.3 | 659  | 4 ABB65280 | Abb65280 Drosophil |
| 43 | 72   | 16.3 | 804  | 4 ABB20282 | Abb20282 Novel hum |
| 44 | 72   | 16.3 | 1013 | 5 AAM50715 | Aam50715 Human TRP |
| 45 | 72   | 16.3 | 1288 | 5 AAE20285 | Aae20285 Human Trp |

ALIGNMENTS

RESULT 1  
ID AAW36006 standard; protein; 156 AA.  
XX AC AAW36006;  
XX AC  
DT 03-MAR-1998 (first entry)  
XX AC  
DE Human Fchd605 gene product.  
XX AC  
KW Fchd605 gene; differential expression; monocyte; human; fcam cell;  
KW cardiovascular disease; atherosclerosis; ischaemia; reperfusion;  
KW hypertension; restenosis; arterial inflammation; therapy; diagnosis;  
KW drug screening; marker.  
XX AC  
OS Homo sapiens.  
XX AC  
PN WO9730065-AL.  
XX AC  
PD 21-AUG-1997.  
XX AC  
PF 14-FEB-1997; 97WO-US002291.  
XX AC  
PR 16-FEB-1996; 96US-0011787P.  
XX AC  
PA 13-FEB-1997; 97US-00799910.  
XX AC  
PI (MILL-) MILLENNIUM PHARM INC.  
XX AC  
PI Falb DA;  
XX AC  
DR WPI; 1997-424966/39.  
XX AC  
DR N-PSDB; AAT94471.  
XX AC  
FT New genes differentially expressed in cardiovascular disease - used for  
FT diagnosis, drug screening and treatment of cardiovascular disease, e.g.  
FT atherosclerosis, restenosis, hypertension, etc.  
XX AC  
PS Example 6; Fig 5; 163pp; English.  
XX AC  
CC This protein is encoded by the novel human fchd605 gene (see AAT94471)  
CC that is up-regulated in monocytes treated with oxidised low density  
CC lipoproteins that stimulate the conditions under which foam cells develop  
CC during atherogenesis. The protein has sequence similarity to the mouse  
CC gly96 gene and to ESR AAT49532. Novel fchd531, fchd540, fchd545, fchd602  
CC and fchd605 genes (see AAT94467-71) provide a fingerprint for the study  
CC of cardiovascular diseases, including atherosclerosis,  
CC ischaemia/reperfusion, hypertension, restenosis and arterial  
CC inflammation. Methods are provided for the diagnosis, monitoring in

CC Clinical trials, screening for therapeutically effective compounds, and  
 CC treatment of cardiovascular diseases based on discoveries regarding the  
 CC expression patterns of these novel genes

XX SQ Sequence 156 AA;

Query Match 100.0%; Score 442; DB 2; Length 156;  
 Best Local Similarity 100.0%; Pred.No. 3.9e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERQLPVEENPAKRLFLILLTIIVFCQILMAEEGVPAFLPEDAPNAASLAPTPVSPVLE 60  
 DE 71 VERQLPVEENPAKRLFLILLTIIVFCQILMAEEGVPAFLPEDAPNAASLAPTPVSPVLE 130

QY 61 PNLITSPSPDYALDLSTFLQHPAAF 86  
 DB 131 PNLITSPSPDYALDLSTFLQHPAAF 156

# RESULT 2

ID RAY45017 standard; protein; 156 AA.

XX AC RAY45017;

XX DT 31-MAY-2000 (first entry)

XX DE Protein encoded by fchd605 gene.

XX KW fchd605 gene; human; cardiovascular disease; oncogenic disorder;  
 KW diabetic retinopathy; fibroproliferative disorder; arteriosclerosis;  
 KW TGF-beta signalling pathway; TGF; Transforming growth factor;  
 KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;  
 KW vascularisation; cytostatic; antidiabetic; ophthalmological.

XX OS Homo sapiens.

XX PN WO200006206-A1.

XX PD 10-FEB-2000.

XX PF 30-JUL-1999; 99WO-US017394.

XX PR 30-JUL-1998; 98US-00126640.

XX PA (MILL-) MILLENNIUM PHARM INC.

XX PI Paib DA;

XX DR WPI; 2000-205414/18.

XX DR N-PSDB; AA250711.

XX Identifying substances for ameliorating symptoms of fibroproliferative  
 XX diseases or oncogenic related disorders.

XX Example; Fig 5; 214pp; English.

XX The patent discloses methods for the treatment and diagnosis of  
 CC cardiovascular diseases by novel human genes which are differentially  
 CC expressed in different cardiovascular disease states. Compositions which  
 CC can modify TGF-beta signalling pathway are identified by screening. These  
 CC are used therapeutically to treat fibroproliferative and oncogenic  
 CC disorders, especially TGF (transforming growth factor)-beta related  
 CC disorders, including diabetic retinopathy, arteriosclerosis, pancreatic  
 CC cancer, angiogenesis, inflammation, fibrosis, tumour growth and  
 CC vascularisation. The present sequence is the protein product of fchd605  
 CC gene which is up-regulated in monocytes treated with oxidised LDL (low  
 CC density lipoprotein) can be used to design cardiovascular disease  
 CC treatment strategies. Depending on whether the up-regulation has a  
 CC pathogenic or protective effect treatment methods can be designed to  
 CC increase or decrease the activity of the protein product of the gene

XX SQ Sequence 156 AA;

Query Match 100.0%; Score 442; DB 3; Length 156;  
 Best Local Similarity 100.0%; Pred.No. 3.9e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VERQLPVEENPAKRLFLILLTIIVFCQILMAEEGVPAFLPEDAPNAASLAPTPVSPVLE 60  
 DE 71 VERQLPVEENPAKRLFLILLTIIVFCQILMAEEGVPAFLPEDAPNAASLAPTPVSPVLE 130

QY 61 PNLITSPSPDYALDLSTFLQHPAAF 86  
 DB 131 PNLITSPSPDYALDLSTFLQHPAAF 156

# RESULT 3

AAU78232

ID AAU78232 standard; protein; 156 AA.

XX AC AAU78232;

XX DT 05-JUN-2002 (first entry)

XX DE Human proliferation regulated gene-1 protein, prg-1/1EX-1.

XX KW Human; MIVR-1; Mechanically Induced Vascular Receptor 1; cytostatic;  
 KW cardiant; cerebroprotective; antiarteriosclerotic; cardiac cell;  
 KW anti-apoptotic; vascular endothelial cell; cardiac hypertrophy;  
 KW myocardial infarction; stroke; arteriosclerosis; heart failure;  
 KW proliferation regulated gene-1; prg-1; 1EX-1.

XX OS Homo sapiens.

XX PN WO200016416-A2.

XX PD 28-FEB-2002.

XX PF 21-AUG-2001; 2001WO-US026089.

XX PR 22-AUG-2000; 2000US-0227159P.

XX PA (BGHM) BRIGHAM & WOMENS HOSPITAL INC.

XX PA (PFIZ) PFIZER INC.

XX PI Lee RT, Landschulz KT, Kennedy SP, Thompson JF, Turi TG;

XX DR WPI; 2002-280912/32.

XX DR N-PSDB; ABK12138.

XX Novel nucleic acid molecule encoding Mechanically Induced Vascular  
 XX Receptor-1 polypeptide, useful for treating cardiovascular diseases.

XX Example; Page 90-91; 105pp; English.

XX The invention relates to an isolated nucleic acid molecule encoding a  
 CC Mechanically Induced Vascular Receptor (MIVR)-1 polypeptide having  
 CC cardiac cell anti-apoptotic activity and fragments of it provided they  
 CC are not identical to Genbank sequences A1761441.1, A1594390, NM 004338  
 CC and A0177461. Also included are expression vectors, host cells, the MIVR-  
 CC 1 polypeptide, MIVR-1 binding peptides, modulators of MIVR-1, contacting  
 CC a molecule having cardiac cell anti-apoptotic activity with a candidate  
 CC agent, where the molecule is a nucleic acid molecule comprising MIVR-1,  
 CC IEX-1, VDUP-1, BTG-2 and TIS-11d or its expression product, determining  
 CC if the anti-apoptotic activity is modulated and thereby identifying a  
 CC modulator. The cardiac cell anti-apoptotic molecules and nucleic acids  
 CC of the invention are useful for treating, diagnosing and monitoring  
 CC progression of such diseases and disorders as characterised by increased  
 CC apoptotic cell-death of vascular endothelial cells e.g. cardiac  
 CC hypertrophy, myocardial infarction, stroke, arteriosclerosis and heart  
 CC failure. The present sequence represents human proliferation regulated  
 CC gene-1, prg-1 (also known as IEX-1) protein which is also mechanically  
 CC induced and has an apoptosis regulatory function

XX SQ Sequence 156 AA;

Query Match 100.0%; Score 442; DB 5; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLFVEEENPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60  
 DB 71 VRRQLFVEEENPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PNLITSEPSDYALDLSTFLQHPAAF 86  
 DB 131 PNLITSEPSDYALDLSTFLQHPAAF 156

RESULT 4  
 ADD46473  
 ID ADD46473 standard; protein; 156 AA.  
 XX  
 AC ADD46473;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human P-otein CAA65304, SEQ ID NO 12154.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GHEO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 PI WPI; 2003-268312/26.  
 DR GENBANK; CAA65304.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating

CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 156 AA;

Query Match 100.0%; Score 442; DB 7; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLFVEEENPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60.  
 DB 71 VRRQLFVEEENPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PNLITSEPSDYALDLSTFLQHPAAF 86  
 DB 131 PNLITSEPSDYALDLSTFLQHPAAF 156

RESULT 5  
 ADD44772  
 ID ADD44772 standard; protein; 156 AA.  
 XX  
 AC ADD44772;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human Protein CAA65304, SEQ ID NO 10203.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GHEO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 XX Woolf C, D'urso D, Befort K, Costigan M;  
 PI WPI; 2003-268312/26.  
 DR GENBANK; CAA65304.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating

expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 156 AA;

Query Match 100.0%; Score 442; DB 7; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.9e-39;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEENPNARKLLFLLLTIVFCQILMAEKGVPALPPEDAPNAASLAPTPVSPVLE 60  
DB 71 VRRQLPVEENPNARKLLFLLLTIVFCQILMAEKGVPALPPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
DB 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 6  
ADD44775  
ID ADD44775 standard; protein; 156 AA.  
XX  
AC ADD44775;  
XX  
XX 29-JAN-2004 (first entry)  
XX  
DE Human Protein CAA65304, SEQ ID NO 10206.  
XX  
XX Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO) GEN HOSPITAL CORP.  
PA (FARB) BAYER AG.  
XX  
XX Woolf C, D'urso D, Befort K, Costigan M;  
XX WPI; 2003-269312/26.  
XX GENBANK; CAA65304.

PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
XX Claim 1; Page; 1017pp; English.

XX The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,

derivative or allelic variation of the nucleic acid sequence. Also claimed are a vector comprising the novel polynucleotide, a host cell comprising the vector, a method for identifying a nucleotide sequence which is differentially regulated in an animal subjected to pain and a kit to perform the method, an array, a method for identifying an agent that increases or decreases the expression of the polynucleotide sequence that is differentially expressed in neuronal tissue of a first animal subjected to pain, a method for identifying a compound which regulates the expression of a polynucleotide sequence which is differentially expressed in an animal subjected to pain, a method for identifying a compound that regulates the activity of one or more of the polynucleotides, a method for producing a pharmaceutical composition, a method for identifying a compound or small molecule that regulates the activity in an animal of one or more of the polypeptides given in the specification, a method for identifying a compound useful in treating pain and a pharmaceutical composition comprising the one or more polypeptides or their antibodies. The polynucleotide or the compound that modulates its activity is useful for preparing a medicament for treating pain (e.g. spinal segmental nerve injury (SNI)), chronic constriction injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene therapy). The sequence presented is a human protein (shown in Table 2 of the specification) which is differentially expressed during pain. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 156 AA;

Query Match 100.0%; Score 442; DB 7; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3.9e-39;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEENPNARKLLFLLLTIVFCQILMAEKGVPALPPEDAPNAASLAPTPVSPVLE 60  
DB 71 VRRQLPVEENPNARKLLFLLLTIVFCQILMAEKGVPALPPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
DB 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 7  
ADD44769  
ID ADD44769 standard; protein; 156 AA.  
XX  
AC ADD44769;  
XX  
XX 29-JAN-2004 (first entry)  
XX  
DE Human Protein CAA65304, SEQ ID NO 10200.  
XX  
XX Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX  
OS Homo sapiens.  
XX  
PN WO2003016475-A2.  
XX  
PD 27-FEB-2003.  
XX  
XX 14-AUG-2002; 2002WO-US025765.  
XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
XX (GEO) GEN HOSPITAL CORP.  
PA (FARB) BAYER AG.

XX Woolf C, D'urso D, Befort K, Costigan M;

XX WPI; 2003-269312/26.

DR GENBANK; CAA65304.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 156 AA;

Query Match 100.0%; Score 442; DB 7; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVSEPNPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60  
 DB 71 VRRQLPVSEPNPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSEPSDYALDLSFTLQHPAAF 86  
 DB 131 PFNLTSEPSDYALDLSFTLQHPAAF 156

RESULT 8  
 ADD44778  
 ID ADD44778 standard; protein; 156 AA.  
 XX  
 AC ADD44778;  
 XX  
 DT 29-JAN-2004 (first entry)  
 DE Human Protein CAA65304, SEQ ID NO 10209.  
 DE  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PF 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX  
 DR WP1; 2003-268112/26.  
 DR GENBANK; CAA65304.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.

CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of  
 CC the specification) which is differentially expressed during pain. Note:  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic form directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 156 AA;

Query Match 100.0%; Score 442; DB 7; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVSEPNPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60  
 DB 71 VRRQLPVSEPNPAKRLIFLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130  
 QY 61 PFNLTSEPSDYALDLSFTLQHPAAF 86  
 DB 131 PFNLTSEPSDYALDLSFTLQHPAAF 156

RESULT 9  
 AAG74873  
 ID AAG74873 standard; protein; 166 AA.  
 XX  
 AC AAG74873;  
 XX  
 DT 03-SEP-2001 (first entry)  
 XX  
 DE Human colon cancer antigen protein SEQ ID NO:5637.  
 XX  
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW colorectal carcinoma; chromosome 6.  
 XX  
 OS Homo sapiens.

XX WC200122920-AA.  
 XX 05-APR-2001.  
 XX 28-SEP-2000; 2000WO-US026524.  
 XX 29-SEP-1999; 99US-0157137P.  
 PR 03-NOV-1999; 99US-0163280P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Ruben SM, Barash SC, Birse CE, Rosen CA;  
 XX WPI: 2001-235357/24.  
 XX N-PSDB; AAH34278.  
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers.  
 XX Claim 11; Page 7188-7189; 9803pp; English.  
 XX AAH32943 to AAH37195 and AAH37514 to AAH37798 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where the  
 CC proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene therapy  
 CC and vaccine production. N and P may be used in the prevention, diagnosis  
 CC and treatment of diseases associated with inappropriate P expression. For  
 CC example, N and P may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of P by expressing inactive proteins or to  
 CC supplement the patients own production of P. Additionally, N may be used  
 CC to produce the colon cancer-associated Ps, by inserting the nucleic acids  
 CC into a host cell and culturing the cell to express the proteins. N and P  
 CC can be used in the prevention, diagnosis and treatment of colorectal  
 CC carcinomas and cancers. AAH37196 to AAH37204 and AAH37789 represent  
 CC sequences used in the exemplification of the present invention. N.B.  
 CC Pages 666 to 682 and page 7053 of the sequence listing were missing at  
 CC time of publication, meaning no sequences are present for SEQ ID NO:1027  
 CC to 1052, 7921 and 7922  
 XX Sequence 166 AA;  
 SQ Query Match 100.0%; Score 442; DB 4; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 4.2e-39;  
 Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 VRRQLVPEPNPAKRLFLLLITVFCQILMAEGVPAIPPEDPAPNAASLAPTPSPVLE 60  
 DB 81 VRRQLVPEPNPAKRLFLLLITVFCQILMAEGVPAIPPEDPAPNAASLAPTPSPVLE 140  
 QY 61 PFNLSEPSDYALDLSTFLQHPAAF 86  
 DB 141 PFNLSEPSDYALDLSTFLQHPAAF 166  
 RESULT 10  
 AA03933  
 ID AA03933 standard; protein; 82 AA.  
 XX AA03933;  
 AC AA03933;  
 XX 09-AUG-2001 (first entry)  
 DT Human gene 36 encoded secreted protein HAUC84, SEQ ID NO:98.  
 DE Human; secreted protein; proliferative disorder; cancer; tumour;  
 XX foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;

KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulnery; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification.  
 XX Homo sapiens.  
 XX Key Location/Qualifiers  
 FH Peptide 1..33  
 FT /label= signal\_peptide  
 FT Protein 34..82  
 FT /note= "Mature secreted protein"  
 XX WO200077022-A1.  
 PN 21-DEC-2000.  
 XX 01-JUN-2000; 2000WO-US015136.  
 XX 11-JUN-1999; 99US-0136629P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM, Komatsoulis GA;  
 XX WPI: 2001-367020/38.  
 XX N-PSDB; AA08380.  
 XX Nucleic acids encoding 50 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating diseases, e.g. Parkinson's  
 PT disease, botulism, cancers and Scimitar syndrome.  
 XX Claim 11; Page 546; 614pp; English.  
 XX AA08345-AA08394 represent cDNAs corresponding to 50 human secreted  
 CC protein genes and AA08388-AA08394 represent the proteins they encode.  
 CC AA08348-AA08396 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 50 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative  
 CC disorders, cancer, tumours, foetal and developmental abnormalities,  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies, disease),  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infectious. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein of the invention  
 XX Sequence 82 AA;  
 SQ Query Match 95.7%; Score 423; DB 4; Length 82;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-37;  
 Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5 LPVEEPNPAKRLFLLLITVFCQILMAEGVPAIPPEDPAPNAASLAPTPSPVLEPFL 64  
 DB 1 LPVEEPNPAKRLFLLLITVFCQILMAEGVPAIPPEDPAPNAASLAPTPSPVLEPFL 50  
 QY 65 TSEPSDYALDLSTFLQHPAAF 86

Db 61 TSEPSDYALDLSTFLQQHPA 82

RESULT 11  
AAE03976

TD AAE03976 standard; peptide; 83 AA.

XX AAE03976;

XX 09-AUG-2001 (first entry)

XX Human gene 36 encoded secreted protein fragment, SEQ ID NO:168.

XX Human; secreted protein; proliferative disorder; cancer; tumour;

KW foetal abnormality; developmental abnormality; haematopoietic disorder;

KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;

KW inflammation; allergy; neurological disorder; Alzheimer's disease;

KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;

KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;

KW cardiovascular disorder; angiogenic disorder; kidney disorder;

KW gastrointestinal disorder; pregnancy-related disorder;

KW endocrine disorder; infection; wound healing; vulnery; cell culture;

KW chemotaxis; food additive; gene therapy; binding partner identification.

XX Homo sapiens.

OS

XX W0200077022-A1.

XX 21-DEC-2000.

XX 01-JUN-2000; 2000WO-US015136.

XX 11-JUN-1999; 99US-0138629P.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Ruben SM, Komatsoulis GA;

XX WPI; 2001-367020/38.

XX Nucleic acids encoding 50 human secreted polypeptides, useful for

PT preventing, diagnosing and/or treating diseases, e.g. Parkinson's

PT disease, botulism, cancers and Scimitar syndrome.

XX Disclosure; Page 591; 61app; English.

XX AAD08345-AA08394 represent cDNAs corresponding to 50 human secreted

CC protein genes and AAE03898-AAE03947 represent the proteins they encode.

CC AAE03948-AAE03996 represent human secreted protein fragments or variants.

CC The genes and their secreted proteins are useful for preventing, treating

CC or ameliorating medical conditions, e.g., by protein or gene therapy.

CC Pathological conditions can be diagnosed by determining the amount of the

CC new protein in a sample or by determining the presence of mutations in

CC the new genes. Specific uses are described for each of the 50 genes,

CC based on the tissues in which they are most highly expressed, and include

CC developing products for the diagnosis or treatment of proliferative

CC disorders, cancer, tumours, foetal and developmental abnormalities,

CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune

CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,

CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),

CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,

CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,

CC angiogenic disorders, kidney disorders, gastrointestinal disorders,

CC pregnancy-related disorders, endocrine disorders, and infections. The

CC proteins can also be used to aid wound healing and epithelial cell

CC proliferation, to prevent skin aging due to sunburn, to maintain organs

CC before transplantation, for supporting cell culture of primary tissues,

CC to regenerate tissues, to identify their cognate ligands or binding

CC partners, and in chemotaxis, and can be used as a food additive or

CC preservative to modify storage properties. Antibodies specific for a

CC protein of the invention can be used in alleviating symptoms associated

CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,

CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The

CC present sequence represents a human secreted protein fragment referred to

CC in the disclosure of the invention

XX

SQ Sequence 83 AA;

Query Match 95.2%; Score 421; DB 4; Length 83;

Best Local Similarity 100.0%; Pred.No. 3.2e-37;

Matches 82; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLFVEEENPAKRLIFLLITVFCQILMAEGVPAPEPPEDAPNAASLAPTPSPVLE 60

Db 2 VRRQLFVEEENPAKRLIFLLITVFCQILMAEGVPAPEPPEDAPNAASLAPTPSPVLE 61

QY 61 PNNLTSEPSDYALDLSTFLQQH 82

Db 62 PNNLTSEPSDYALDLSTFLQQH 83

RESULT 12  
AAG75952

ID AAG75952 standard; protein; 83 AA.

XX AAG75952;

XX 03-SEP-2001 (first entry)

DT Human colon cancer antigen protein SEQ ID NO:6716.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;

KW colorectal carcinoma; chromosome 6.

XX Homo sapiens.

XX W0200122920-A2.

XX 05-APR-2001.

XX 28-SEP-2000; 2000WO-US026524.

XX 29-SEP-1999; 99US-0157137P.

PR 03-NOV-1999; 99US-0163280P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Barash SC, Birse CE, Rosen CA;

PI WPI; 2001-235357/24.

DR N-PSDE; AAH35357.

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,

PT useful for preventing, diagnosing and/or treating colorectal cancers.

XX Claim 11; Page 8177-8178; 9803pp; English.

XX AAH32943 to AAH37195 and AAG7788 represent human colon

CC cancer-associated nucleic acid molecules (N) and proteins (P), where the

CC proteins are collectively known as colon cancer antigens. The colon

CC cancer antigens have cytostatic activity and can be used in gene therapy

CC and vaccine production. N and P may be used in the prevention, diagnosis

CC and treatment of diseases associated with inappropriate P expression. For

CC example, N and P may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of P by expressing inactive proteins or to

CC supplement the patients own production of P. Additionally, N may be used

CC to produce the colon cancer-associated Ps, by inserting the nucleic acids

CC into a host cell and culturing the cell to express the proteins. N and P

CC can be used in the prevention, diagnosis and treatment of colorectal

CC carcinomas and cancers. AAH37196 to AAH37204 and AAH7789 represent

CC sequences used in the exemplification of the present invention. N.B.

CC Pages 666 to 682 and page 7053 of the sequence listing were missing at

CC time of publication, meaning no sequences are present for SEQ ID NO:1027

CC to 1052, 7921 and 7922



XX SQ Sequence 83 AA;  
 Query Match 58.6%; Score 259; DB 4; Length 83;  
 Best Local Similarity 89.5%; Pred. No. 5.7e-20;  
 Matches 51; Conservative 0; Mismatches 6; Indels 0; Gaps 0;  
 QY 30 MAEGVPAPEPDAAPNAASLAPTPVSPVLEPFLNLTSEPSDYALDLSFTLQHPAAF 86  
 DB 2 MAEGXPAPLPEDAPNAASLAPTPKSPLEPFLNLTSEPSDXALDLSFTLQXPDFAF 58  
 RESULT 13  
 AAEL1773  
 ID AAEL1773 standard; protein; 923 AA.  
 AC AAEL1773;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human kinase (PKIN)-7 protein.  
 XX  
 KW Human kinase; PKIN; gene therapy; adenocarcinoma; immune disorder; gout;  
 KW cancer; allergy; sarcoma; leukaemia; acquired immune deficiency syndrome;  
 KW AIDS; Addison's disease; microbial infection; inflammation; osteoporosis;  
 KW atherosclerosis; cardiovascular disease; myocardial infarction; anaemia;  
 KW myasthenia gravis; cataract; growth and development disorder;  
 KW seizure disorder; pulmonary embolism; Gaucher's disease; lipid disorder;  
 KW lipid storage disease; Pick's disease; Tay-Sachs disease; renal disease;  
 KW obesity; restorative therapy; immunomodulatory; vaccine; cardiovascular;  
 KW antimicrobial; cytostatic; antiinflammatory; asthma; chromosome 11.  
 XX  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 24..268  
 FT /note= "Eukaryotic protein kinase domain"  
 FT Domain 25..259  
 FT /note= "Protein kinase domain"  
 FT Region 93..106  
 FT /note= "Tyrosine kinase catalytic site"  
 FT Region 129..147  
 FT /note= "Tyrosine kinase catalytic site"  
 FT Region 195..217  
 FT /note= "Tyrosine kinase catalytic site"  
 XX  
 WO200181555-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US012992.  
 XX  
 PR 20-APR-2000; 2000US-0199021P.  
 PR 28-APR-2000; 2000US-0200226P.  
 PR 05-MAY-2000; 2000US-0202339P.  
 PR 11-MAY-2000; 2000US-0203505P.  
 PR 18-MAY-2000; 2000US-0205564P.  
 PR 26-MAY-2000; 2000US-0207739P.  
 PR 01-JUN-2000; 2000US-0208795P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Yue H, Gandhi AR, Tribouley CM, Kearney L, Griffin JA, Nguyen DB;  
 PI Bandman O, Lu DM, Lal P, Burford N, Khan FA, Walla NK, Yao MG;  
 PI Patterson C, Burrill JD, Marcus GA, Zingler KA, Recipon SA, Lu Y;  
 PI Policky JV, Thornton M, Tang YT, Hafalia A, Elliott VS, Baughn MR;  
 PI Walsh RT, Ramkumar J, Borowsky ML, Au-Young J, Hillman JL;  
 PI Gururajan R;  
 XX  
 DR WFI; 2001-611740/70.  
 DR N-PSDB; AAD18822.  
 XX  
 PI Human kinases and nucleic acids, useful for preventing diagnosing and

PT treating cancers, inflammation and immune disorders.  
 XX  
 PS Claim 1; Page 131-133; 166pp; English.  
 XX  
 CC The present invention relates to human kinases (PKIN) and the nucleic  
 CC acids encoding them. PKIN is used as vaccine and in gene therapy. PKIN is  
 CC used in the prevention, diagnosis and treatment of diseases cancers,  
 CC adenocarcinoma, leukaemia, sarcoma, immune disorder, Addison's disease,  
 CC acquired immune deficiency syndrome (AIDS), anaemia, asthma, allergies,  
 CC gout, microbial infections, cardiovascular disease and/or inflammation,  
 CC myasthenia gravis, atherosclerosis, cirrhosis, osteoporosis, myocardial  
 CC infarction, cataract, growth and development disorder, seizure disorder,  
 CC pulmonary embolism, Gaucher's disease, lipid disorder, lipid storage  
 CC disease, Pick's disease, Tay-Sachs disease, renal disease and obesity.  
 CC PKIN may be used to treat disorders associated with decreased PKIN  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of PKIN by expressing inactive proteins or to  
 CC supplement the patient's own production of PKIN. PKIN nucleic acids may be  
 CC used to produce the PKIN polypeptide, by inserting the nucleic acids into  
 CC a host cell and culturing the cell to express the protein. PKIN nucleic  
 CC acid and its complementary sequences may also be used as DNA probes in  
 CC diagnostic assays to detect and quantitate the presence of similar  
 CC nucleic acid sequences in samples and therefore which patients may be in  
 CC need of restorative therapy. The present sequence is human PKIN-7  
 CC protein. The PKIN-7 sequence was mapped to chromosome 11 within the  
 CC interval from 104.8 to 117.9 centiMorgans  
 XX  
 SQ Sequence 923 AA;  
 Query Match 19.1%; Score 84.5; DB 4; Length 923;  
 Best Local Similarity 33.3%; Pred. No. 3.8;  
 Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;  
 QY 7 VEEPNPAKGLLEL-----LLTIVFCQILMAEGVP-----APLPEDAPNAASLAP----- 52  
 DB 709 LQEHRLQCKRLFLQKOSQLQAYENQMOTABSSYPQPSQQLPLPQETPFPQQAPPFSILT 768  
 QY 53 TPVSPVLEPFLNLTSEPSDYALDLSFTLQOH 82  
 DB 769 QPLSPVLEP-----SSEQMQV-----SFFLSQY 791  
 RESULT 14  
 AA865631  
 ID AA865631 standard; protein; 926 AA.  
 XX  
 AC AA865631;  
 XX  
 DT 27-MAR-2001 (first entry)  
 XX  
 DE Novel protein kinase, SEQ ID NO: 158.  
 XX  
 KW Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;  
 KW immunosuppressive; cardiac; renal; antiinflammatory; antiasthmatic;  
 KW dermatological; antidiabetic; antifertility; gene therapy; vaccine;  
 KW immune disorder; cardiovascular disease; neurodegenerative disease;  
 KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;  
 KW inflammatory pelvic disease; multiple sclerosis; psoriasis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200073469-A2.  
 XX  
 PD 07-DEC-2000.  
 XX  
 PF 26-MAY-2000; 2000WO-US014842.  
 XX  
 PR 28-MAY-1999; 99US-0136503P.  
 XX  
 PA (SUGEN-) SUGEN INC.  
 XX  
 PI Flowman GD, Martinez R, Whyte D, Sudersanam S;  
 XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:33:42 ; Search time 23 Seconds  
(without alignments)  
193.036 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156  
Perfect score: 442  
Sequence: 1 VRQPLPVEPNPKAKLLFL.....PSPDYALDLSFLQHPAAP 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA: \*  
1: /cgn2\_6/prodata/2/iaa/5A COMB.pep: \*  
2: /cgn2\_6/prodata/2/iaa/5B COMB.pep: \*  
3: /cgn2\_6/prodata/2/iaa/6A COMB.pep: \*  
4: /cgn2\_6/prodata/2/iaa/6B COMB.pep: \*  
5: /cgn2\_6/prodata/2/iaa/PCTUS COMB.pep: \*  
6: /cgn2\_6/prodata/2/iaa/backfiles1.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description          |
|------------|-------|-------------|--------|----|----------------------|
| 1          | 442   | 100.0       | 156    | 3  | US-08-826-246-10     |
| 2          | 442   | 100.0       | 156    | 3  | US-08-944-495-10     |
| 3          | 442   | 100.0       | 156    | 3  | US-09-126-640-11     |
| 4          | 442   | 100.0       | 156    | 3  | US-08-925-588-10     |
| 5          | 442   | 100.0       | 156    | 4  | US-09-288-292A-11    |
| 6          | 442   | 100.0       | 156    | 4  | US-09-372-044-10     |
| 7          | 442   | 100.0       | 156    | 4  | US-08-825-486-10     |
| 8          | 79.5  | 18.0        | 1596   | 3  | US-09-356-952-3      |
| 9          | 76.5  | 17.3        | 1843   | 3  | US-09-413-814-50     |
| 10         | 72.5  | 16.4        | 244    | 4  | US-09-252-991A-22431 |
| 11         | 71.5  | 16.2        | 610    | 1  | US-07-821-717B-6     |
| 12         | 71.5  | 16.2        | 610    | 1  | US-08-119-262B-6     |
| 13         | 71.5  | 16.2        | 610    | 1  | US-08-135-929A-11    |
| 14         | 71.5  | 16.2        | 610    | 1  | US-08-234-265A-11    |
| 15         | 70    | 15.8        | 2972   | 4  | US-09-579-181-2      |
| 16         | 70    | 15.8        | 3118   | 4  | US-09-579-181-1      |
| 17         | 69.5  | 15.7        | 739    | 3  | US-09-035-648-24     |
| 18         | 69.5  | 15.7        | 739    | 3  | US-09-001-951-24     |
| 19         | 69.5  | 15.7        | 739    | 4  | US-08-818-829-24     |
| 20         | 68.5  | 15.5        | 204    | 4  | US-08-529-055-58     |
| 21         | 68.5  | 15.5        | 878    | 1  | US-08-237-919-2      |
| 22         | 68.5  | 15.5        | 878    | 4  | US-08-732-429-2      |
| 23         | 68.5  | 15.5        | 878    | 4  | US-09-798-267-2      |
| 24         | 68.5  | 15.5        | 878    | 4  | US-09-798-267-3      |
| 25         | 68.5  | 15.5        | 878    | 5  | PCT-US95-05518-2     |
| 26         | 68.5  | 15.5        | 1337   | 3  | US-08-854-585-2      |
| 27         | 68.5  | 15.5        | 1337   | 4  | US-09-447-533-2      |

|    |      |      |      |   |                      |
|----|------|------|------|---|----------------------|
| 28 | 68.5 | 15.5 | 1337 | 5 | PCT-US95-05512-2     |
| 29 | 68.5 | 15.5 | 8991 | 4 | US-08-714-741-32     |
| 30 | 68   | 15.4 | 1560 | 4 | US-09-264-512B-2     |
| 31 | 66.5 | 15.0 | 544  | 4 | US-09-489-039A-7235  |
| 32 | 66   | 14.9 | 411  | 4 | US-09-252-991A-18602 |
| 33 | 66   | 14.9 | 2509 | 4 | US-09-252-991A-16642 |
| 34 | 65.5 | 14.8 | 128  | 4 | US-08-529-055-57     |
| 35 | 65.5 | 14.8 | 286  | 1 | US-08-382-184-3      |
| 36 | 65.5 | 14.8 | 286  | 2 | US-08-641-356-3      |
| 37 | 65.5 | 14.8 | 286  | 3 | US-09-132-528-4      |
| 38 | 65.5 | 14.8 | 286  | 3 | US-08-875-494-3      |
| 39 | 65.5 | 14.8 | 286  | 4 | US-09-599-366-4      |
| 40 | 65.5 | 14.8 | 286  | 4 | US-08-875-494-3      |
| 41 | 65.5 | 14.8 | 286  | 4 | US-09-510-031A-7     |
| 42 | 65.5 | 14.8 | 286  | 4 | US-09-985-372-4      |
| 43 | 65.5 | 14.8 | 325  | 1 | US-08-382-184-2      |
| 44 | 65.5 | 14.8 | 325  | 2 | US-08-641-356-2      |
| 45 | 65.5 | 14.8 | 325  | 3 | US-09-132-528-2      |

ALIGNMENTS

RESULT 1  
US-08-826-246-10  
; Sequence 10, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cortuzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 156 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: internal  
US-08-826-246-10

```
Query Match      100.0%; Score 442; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEENPAKRLFLLLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
DB 71 VRRQLPVEENPAKRLFLLLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PENLTSEPSDYALDLSTFLQHPAAAF 86
DB 131 PFNLTSEPSDYALDLSTFLQHPAAAF 156

RESULT 2
US-08-944-495-10
; Sequence 10, Application US/08944495
; Patent No. 6087477
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,495
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
; TELEFAX: (212)8699741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-944-495-10

Query Match      100.0%; Score 442; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEENPAKRLFLLLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
DB 71 VRRQLPVEENPAKRLFLLLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PENLTSEPSDYALDLSTFLQHPAAAF 86
DB 131 PFNLTSEPSDYALDLSTFLQHPAAAF 156
```

```
RESULT 3
US-09-126-640-11
; Sequence 11, Application US/09126640A
; Patent No. 6099823
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; FILE REFERENCE: 7853-126
; CURRENT APPLICATION NUMBER: US/09/126.640A
; CURRENT FILING DATE: 1998-07-30
; EARLIER APPLICATION NUMBER: 08/870,434
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 08/799,910
; EARLIER FILING DATE: 1997-02-13
; EARLIER APPLICATION NUMBER: 60/011,787
; EARLIER FILING DATE: 1996-02-16
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-126-640-11

Query Match      100.0%; Score 442; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEENPAKRLFLLLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
DB 71 VRRQLPVEENPAKRLFLLLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSEPSDYALDLSTFLQHPAAAF 86
DB 131 PFNLTSEPSDYALDLSTFLQHPAAAF 156

RESULT 4
US-08-925-588-10
; Sequence 10, Application US/08925588
; Patent No. 6221628
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-SEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/925,588
; FILING DATE: 08-Sep-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-067-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)7909090
```

```
/
/ TELEFAX: (212)8699741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 156 amino acids
/   TYPE: amino acid
/   STRANDEDNESS: unknown
/   TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ FRAGMENT TYPE: internal
/ SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-08-925-588-10

Query Match      100.0%; Score 442; DB 3; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PPNLTSEPSDYALDLSTFLQHPAAF 86
Db 131 PPNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 5
US-09-288-292A-11
/ Sequence 11, Application US/09288292A
/ Patent No. 6359194
/ GENERAL INFORMATION:
/ APPLICANT: Dean A. Falb
/ APPLICANT: Katherine Galvin
/ APPLICANT: Michael Donovan
/ APPLICANT: Dennis Huszar
/ APPLICANT: Michael A. Gimbrone, Jr.
/ TITLE OF INVENTION: Compositions and Methods for the Treatment and Diagnosis of
/ TITLE OF INVENTION: Cardiovascular Disease
/ FILE REFERENCE: 7853-140-999
/ CURRENT APPLICATION NUMBER: US/09/288,292A
/ PRIOR FILING DATE: 1999-04-08
/ PRIOR APPLICATION NUMBER: 08/870,434
/ PRIOR FILING DATE: 1997-06-06
/ PRIOR APPLICATION NUMBER: 08/799,910
/ PRIOR FILING DATE: 1997-02-13
/ PRIOR APPLICATION NUMBER: 60/011,787
/ PRIOR FILING DATE: 1996-02-16
/ PRIOR APPLICATION NUMBER: 08/485,573
/ PRIOR FILING DATE: 1995-06-07
/ PRIOR APPLICATION NUMBER: 08/386,844
/ PRIOR FILING DATE: 1995-02-10
/ NUMBER OF SEQ ID NOS: 46
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11
/ LENGTH: 156
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-288-292A-11

Query Match      100.0%; Score 442; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PPNLTSEPSDYALDLSTFLQHPAAF 86
Db 131 PPNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 6
```

```
US-09-372-044-10
/ Sequence 10, Application US/09372044A
/ Patent No. 6492126
/ GENERAL INFORMATION:
/ APPLICANT: Dean Falb et al.
/ TITLE OF INVENTION: Compositions and Methods for the
/ TITLE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease
/ FILE REFERENCE: 7853-152
/ CURRENT APPLICATION NUMBER: US/09/372,044A
/ CURRENT FILING DATE: 1999-08-11
/ NUMBER OF SEQ ID NOS: 44
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10
/ LENGTH: 156
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-372-044-10

Query Match      100.0%; Score 442; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRQLPVEEPNPAKRLFLLLTIIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PPNLTSEPSDYALDLSTFLQHPAAF 86
Db 131 PPNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 7
US-08-825-486-10
/ Sequence 10, Application US/08825486
/ Patent No. 6534641
/ GENERAL INFORMATION:
/ APPLICANT: Falb, Dean
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
/ TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF
/ TITLE OF INVENTION: CARDIOVASCULAR DISEASE
/ NUMBER OF SEQUENCES: 44
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: PENNIE & EDMONDS LLP
/ STREET: 1155 Avenue of the Americas
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10036-2711
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/825,486
/ FILING DATE: 28-MAR-1997
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 08/799,910
/ FILING DATE: 13-FEB-1997
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Coruzzi, Laura A
/ REGISTRATION NUMBER: 30,742
/ REFERENCE/DOCKET NUMBER: 7853-077-999
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212)7909090
/ TELEFAX: (212)8699741
/ TELEX: 66141 PENNIE
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/   LENGTH: 156 amino acids
/   TYPE: amino acid
/   STRANDEDNESS: unknown
/
```

```
;
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-825-486-10

Query Match 100.0%; Score 442; DB 4; Length 156;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRLPVEENPAKLLFLLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
DB 71 VRRLPVEENPAKLLFLLLITIVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSPSYALDLSFLQHPAAF 86
DB 131 PFNLTSPSYALDLSFLQHPAAF 156

RESULT 8
US-09-356-952-3
; Sequence 3, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogi, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; EARLIER FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-356-952-3

Query Match 18.0%; Score 79.5; DB 3; Length 1596;
Best Local Similarity 28.3%; Pred. No. 2.5;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRRLPVE-----EPNAKLLFLLLITIVFCQILMAEGVPAPLPEDAPNAASLAP 53
DB 1423 IRENSAIEKRAAATSQNQAAAGPISITLVTSQVATDEPLDFPISF-----AASSST 1477

QY 54 --PVGFLVEPF--NLTSPE-----SDYALDLSFLQ-----HPAAF 86
DB 1478 TSLIFAMSPSPNIPSHFVSTSSSYAHQLMRQQQQQQTTPAIY 1523

RESULT 9
US-09-413-814-50
; Sequence 50, Application US/09413814
; Patent No. 6225064
; GENERAL INFORMATION:
; APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
; APPLICANT: Bristol-Myers Squibb, Co.
; APPLICANT: Beyer, Stefan
; APPLICANT: Bloeker, Helmut
; APPLICANT: Brandt, Petra
; APPLICANT: Cinc, Paul M
; APPLICANT: Dougherty, Brian A
; APPLICANT: Goldberg, Steven L
; APPLICANT: Hofle, Gerhard
; APPLICANT: Mueller, Joachim
; APPLICANT: Reichenbach, Hans
; TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
```

```
;
; TITLE OF INVENTION: heteropolyketide compounds
; FILE REFERENCE: PCT/US 99/23535
; CURRENT APPLICATION NUMBER: US/09/413,814
; CURRENT FILING DATE: 1999-10-07
; EARLIER APPLICATION NUMBER: DE 198 46 493.2
; EARLIER FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 50
; LENGTH: 1843
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-413-814-50

Query Match 17.3%; Score 76.5; DB 3; Length 1843;
Best Local Similarity 35.6%; Pred. No. 6.5;
Matches 21; Conservative 9; Mismatches 24; Indels 5; Gaps 3;

QY 28 ILMAEGVPAPLPEDAPNAASLAPTPVSPVLEPNLTSPSYALDLSFLQHPAAF 86
DB 1443 VATASESQSLPQ--PSTAYATSPQSP--SPVVTG-PADLPAILAAYGQNPHEF 1496

RESULT 10
US-09-252-991A-22431
; Sequence 22431, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 22431
; LENGTH: 244
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22431

Query Match 16.4%; Score 72.5; DB 4; Length 244;
Best Local Similarity 35.1%; Pred. No. 1.5;
Matches 20; Conservative 10; Mismatches 22; Indels 5; Gaps 2;

QY 15 RLFLFLITVFCQILMAEGVPAPLP--PEDAPNAASLAPTPVSPVLEPNLTSEP 68
DB 135 RVLFLLALVLAG--LQPTLAPLPDELPAALPAFFVAPSPMGSAPOQPEAEP 189

RESULT 11
US-07-821-717B-6
; Sequence 6, Application US/07821717B
; Patent No. 5298239
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Cunningham, David
; APPLICANT: Lyle, Vicki A.
; APPLICANT: Fisch, Clara A.
; TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
; TITLE OF INVENTION: GLYCOPROTEIN ID ALPHA LESS REACTIVE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/821,717B
; FILING DATE: 15-JAN-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Timain, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20884/21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PUBLICATION INFORMATION:
; AUTHORS: Lopez, Jose A.
; AUTHORS: Chung, Dominic W.
; AUTHORS: Fujikawa, Kazuo
; AUTHORS: Hagen, Frederick S.
; AUTHORS: Papayannopoulou, Thalia
; AUTHORS: Roth, Gerald J.
; TITLE: Cloning of the alpha chain of human
; TITLE: Platelet glycoprotein Ib: A transmembrane protein with homology
; TITLE: to leucine-rich alpha-2-glycoprotein
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 84
; PAGES: 5615-5619
; DATE: AUG-1987
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610
; PUBLICATION INFORMATION:
; AUTHORS: Zimmerman, Theodore S.
; AUTHORS: Ruggeri, Zaverio M.
; AUTHORS: Houghten, Richard A.
; AUTHORS: Vincete, Vincete
; AUTHORS: Mohri, Hiroshi
; TITLE: Proteolytic fragments and synthetic
; TITLE: peptides that block the binding of von Willebrand factor to the
; TITLE: platelet membrane glycoprotein Ib
; DOCUMENT NUMBER: EP 0 317 278 A2
; FILING DATE: 16-NOV-1988
; PUBLICATION DATE: 24-MAY-1989
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
;
; US-07-821-717B-6
;
; Query Match 16.2%; Score 71.5; DB 1; Length 610;
; Best Local Similarity 45.0%; Pred. No. 6.1;
; Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;
;
; QY 36 PAPLP-----PEDAPNASLAPTVSPVLEPFNLTSPS 69
; | | | | | | | | | | | | | | | | | | | |
; | | | | | | | | | | | | | | | | | | | |
; Db 366 PTPSPTTSEFPVPEPAPNMTLEPTSPPTPEP---TSEPA 402
;
; RESULT 12
; US-08-119-262B-6
; Sequence 6, Application US/081:9262B
; Patent No. 5492809
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Cunningham, David
; APPLICANT: Lyle, Vicki A.
; APPLICANT: Finch, Clara N.
; TITLE OF INVENTION: MUTATIONS RENDERING PLATELET
; TITLE OF INVENTION: GLYCOPROTEIN IB ALPHA LESS REACTIVE
;
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,262B
; FILING DATE: 09-SEP-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/821,717
; FILING DATE: 15-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Timain, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20884/22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PUBLICATION INFORMATION:
; AUTHORS: Lopez, Jose A.
; AUTHORS: Chung, Dominic W.
; AUTHORS: Fujikawa, Kazuo
; AUTHORS: Hagen, Frederick S.
; AUTHORS: Papayannopoulou, Thalia
; AUTHORS: Roth, Gerald J.
; TITLE: Cloning of the alpha chain of human
; TITLE: glycoprotein Ib: A transmembrane protein
; TITLE: leucine-rich alpha-2-glycoprotein
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 84
; PAGES: 5615-5619
; DATE: AUG-1987
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 610
; PUBLICATION INFORMATION:
; AUTHORS: Zimmerman, Theodore S.
; AUTHORS: Ruggeri, Zaverio M.
; AUTHORS: Houghten, Richard A.
; AUTHORS: Vincete, Vincete
; AUTHORS: Mohri, Hiroshi
; TITLE: Proteolytic fragments and synthetic
; TITLE: peptides that block the binding of von Willebrand
; TITLE: membrane glycoprotein Ib
; DOCUMENT NUMBER: EP 0 317 278 A2
; FILING DATE: 16-NOV-1988
; PUBLICATION DATE: 24-MAY-1989
; RELEVANT RESIDUES IN SEQ ID NO: 6: FROM 0 TO 293
;
; US-08-119-262B-6
;
; Query Match 16.2%; Score 71.5; DB 1; Length 610;
; Best Local Similarity 45.0%; Pred. No. 6.1;
; Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;
;
; QY 36 PAPLP-----PEDAPNASLAPTVSPVLEPFNLTSPS 69
; | | | | | | | | | | | | | | | | | | | |
; | | | | | | | | | | | | | | | | | | | |
; Db 366 PTPSPTTSEFPVPEPAPNMTLEPTSPPTPEP---TSEPA 402
;
; factor to the

```

```

RESULT 13
US-08-135-929A-11
; Sequence 11, Application US/08135929A
; Patent No. 593959
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Cunningham, David
; APPLICANT: Lyle, Vicki A.
; APPLICANT: Finch, Clara N.
; APPLICANT: Pincus, Matthew R.
; TITLE OF INVENTION: Mutations in the Gene Encoding the Alpha
; TITLE OF INVENTION: Chain of Platelet Glycoprotein Ib
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/135,929A
; FILING DATE: 14-OCT-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20884/23
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; TELEX: 978450
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-135-929A-11

Query Match 16.2%; Score 71.5; DB 1; Length 610;
Best Local Similarity 45.0%; Pred. No. 6.1;
Matches 18; Conservative 2; Mismatches 11; Indels 9; Gaps 2;

QY 36 PAPLP-----PEDAPNAASLAPTVSPVLEPFLNLTSEPS 69
DB 366 PTFPTTSEPVPEPAPNMTLEPTFTPTTPEP---TSEPA 402

RESULT 14
US-08-234-265A-11
; Sequence 11, Application US/08234265A
; Patent No. 5624817
; GENERAL INFORMATION:
; APPLICANT: Miller, Jonathan L.
; APPLICANT: Cunningham, David
; APPLICANT: Lyle, Vicki A.
; APPLICANT: Finch, Clara N.
; APPLICANT: Pincus, Matthew R.
; TITLE OF INVENTION: Mutations in the Gene Encoding the
; TITLE OF INVENTION: Chain of Platelet Glycoprotein Ib
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/234,265A
; FILING DATE: 28-APR-1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Timian, Susan J.
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 20884/24
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1636
; TELEFAX: (716) 263-1600
; TELEX: 978450
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 610 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-234-265A-11

Query Match 15.8%; Score 70; DB 4; Length 2972;
Best Local Similarity 36.8%; Pred. No. 63;
Matches 21; Conservative 3; Mismatches 23; Indels 10; Gaps 2;

QY 10 PNPAKELLFLGLTTVFCQILMAEEGVPAIPPEPAPNAASLAPTVS-----PYLEP 61
DB 1304 PNPAPQASLL-----APASSASQALATPLAPMAAPOTAILAPSPAPPLAPLFVYAP 1355

```

Search completed: February 25, 2004, 05:37:27  
Job time : 40 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:36:08 ; Search time 34 Seconds  
(without alignments)  
534.094 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442  
Sequence: 1 VRQLVVEEPNPKRLFL.....FPSYALDLSTFLOQHPAA 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 21153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/prodata/1/pubaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/prodata/1/pubaa/US07\_PUBCOMB.pep.\*  
3: /cgn2\_6/prodata/1/pubaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/prodata/1/pubaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/prodata/1/pubaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/prodata/1/pubaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/prodata/1/pubaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/prodata/1/pubaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/prodata/1/pubaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/prodata/1/pubaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/prodata/1/pubaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/prodata/1/pubaa/US09C\_NEW\_PUB.pep.\*  
13: /cgn2\_6/prodata/1/pubaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/prodata/1/pubaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/prodata/1/pubaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/prodata/1/pubaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/prodata/1/pubaa/US10E\_NEW\_PUB.pep.\*  
18: /cgn2\_6/prodata/1/pubaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID | Description         |
|------------|-------|---------------|--------|----|---------------------|
| 1          | 442   | 100.0         | 156    | 8  | US-08-825-486-10    |
| 2          | 442   | 100.0         | 156    | 8  | US-08-870-434-11    |
| 3          | 442   | 100.0         | 156    | 9  | US-09-372-044-10    |
| 4          | 442   | 100.0         | 156    | 9  | US-09-934-249-5     |
| 5          | 442   | 100.0         | 156    | 10 | US-09-560-150-11    |
| 6          | 442   | 100.0         | 156    | 14 | US-10-067-741-11    |
| 7          | 442   | 100.0         | 166    | 14 | US-10-106-698-5647  |
| 8          | 259   | 58.6          | 83     | 14 | US-10-106-698-6726  |
| 9          | 84.5  | 19.1          | 923    | 15 | US-10-258-106-7     |
| 10         | 78    | 17.6          | 5008   | 15 | US-10-051-874-166   |
| 11         | 78    | 17.6          | 5159   | 15 | US-10-085-198-112   |
| 12         | 78    | 17.6          | 5262   | 15 | US-10-051-874-165   |
| 13         | 75.5  | 17.1          | 242    | 9  | US-09-864-761-36180 |
| 14         | 75.5  | 17.1          | 434    | 10 | US-09-298-523B-8    |
| 15         | 72.5  | 16.4          | 111    | 9  | US-09-864-761-39500 |

|    |      |      |      |    |                     |                   |
|----|------|------|------|----|---------------------|-------------------|
| 16 | 72   | 16.3 | 1013 | 9  | US-09-891-762-5     | Sequence 5, Appli |
| 17 | 72   | 16.3 | 1013 | 15 | US-10-391-399-35    | Sequence 25, Appl |
| 18 | 72   | 16.3 | 5262 | 15 | US-10-051-874-167   | Sequence 167, App |
| 19 | 71.5 | 16.2 | 189  | 9  | US-09-810-264-20    | Sequence 20, Appl |
| 20 | 71   | 16.1 | 740  | 15 | US-10-369-493-3700  | Sequence 3700, Ap |
| 21 | 71   | 16.1 | 1000 | 14 | US-10-128-714-3305  | Sequence 3305, Ap |
| 22 | 70.5 | 16.0 | 430  | 14 | US-10-029-386-32401 | Sequence 32401, A |
| 23 | 70   | 15.8 | 434  | 15 | US-10-108-260A-3595 | Sequence 3595, Ap |
| 24 | 70   | 15.8 | 2971 | 14 | US-10-146-473-50    | Sequence 50, Appl |
| 25 | 69.5 | 15.7 | 118  | 9  | US-09-864-761-41420 | Sequence 41420, A |
| 26 | 69.5 | 15.7 | 677  | 14 | US-10-103-313-384   | Sequence 384, App |
| 27 | 69.5 | 15.7 | 843  | 14 | US-10-205-823-401   | Sequence 401, App |
| 28 | 69.5 | 15.7 | 902  | 9  | US-09-922-199A-2    | Sequence 2, Appli |
| 29 | 69.5 | 15.7 | 1286 | 14 | US-10-156-761-14323 | Sequence 14323, A |
| 30 | 69.5 | 15.7 | 1887 | 15 | US-10-369-493-22455 | Sequence 22455, A |
| 31 | 69   | 15.6 | 1212 | 15 | US-10-369-493-5414  | Sequence 5414, A  |
| 32 | 69   | 15.6 | 1454 | 15 | US-10-369-493-5413  | Sequence 5413, Ap |
| 33 | 68.5 | 15.5 | 191  | 9  | US-09-950-933A-66   | Sequence 66, Appl |
| 34 | 68.5 | 15.5 | 194  | 9  | US-09-764-864-1027  | Sequence 1027, Ap |
| 35 | 68.5 | 15.5 | 194  | 9  | US-09-764-864-1456  | Sequence 1456, Ap |
| 36 | 68.5 | 15.5 | 370  | 10 | US-09-946-274-315   | Sequence 315, App |
| 37 | 68.5 | 15.5 | 370  | 13 | US-10-052-586-350   | Sequence 350, App |
| 38 | 68.5 | 15.5 | 370  | 14 | US-10-174-590-350   | Sequence 350, App |
| 39 | 68.5 | 15.5 | 370  | 14 | US-10-176-758-350   | Sequence 350, App |
| 40 | 68.5 | 15.5 | 370  | 14 | US-10-175-737-350   | Sequence 350, App |
| 41 | 68.5 | 15.5 | 370  | 14 | US-10-173-706-350   | Sequence 350, App |
| 42 | 68.5 | 15.5 | 370  | 14 | US-10-175-738-350   | Sequence 350, App |
| 43 | 68.5 | 15.5 | 370  | 14 | US-10-175-752-350   | Sequence 350, App |
| 44 | 68.5 | 15.5 | 370  | 14 | US-10-176-482-350   | Sequence 350, App |
| 45 | 68.5 | 15.5 | 370  | 14 | US-10-176-757-350   | Sequence 350, App |

ALIGNMENTS

RESULT 1  
US-08-825-486-10  
; Sequence 10, Application US/08825486  
; Publication No. US20020016303A1  
; GENERAL INFORMATION:  
; APPLICANT: Faltb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/825,486  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-077-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)969090  
; TELEFAX: (212)9699741  
; INFORMATION FOR SEQ ID NO: 10:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-825-486-10

Query Match      100.0%; Score 442; DB 8; Length 156;
Best Local Similarity 100.0%; Pred. NO. 3e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSPSDYALDLSTFLQHPAAF 86
Db 131 PFNLTSPSDYALDLSTFLQHPAAF 156

RESULT 2
US-08-870-434-11
; Sequence 11, Application US/08870434
; Publication No. US20020034736A1
; GENERAL INFORMATION:
; APPLICANT: Falb, Dean
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITL OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/870,434
; FILING DATE: 06-JUN-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/799,910
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7853-084
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-870-434-11

Query Match      100.0%; Score 442; DB 8; Length 156;
Best Local Similarity 100.0%; Pred. NO. 3e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSPSDYALDLSTFLQHPAAF 86
Db 131 PFNLTSPSDYALDLSTFLQHPAAF 156

```

```

Db 71 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130
QY 61 PFNLTSPSDYALDLSTFLQHPAAF 86
Db 131 PFNLTSPSDYALDLSTFLQHPAAF 156

RESULT 3
US-09-372-044-10
; Sequence 10, Application US/09372044A
; Patent No. US20020102603A1
; GENERAL INFORMATION:
; APPLICANT: Dean Falb et al.
; TITLE OF INVENTION: Compositions and Methods for the
; TITL OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease
; FILE REFERENCE: 7853-152
; CURRENT APPLICATION NUMBER: US/09/372,044A
; CURRENT FILING DATE: 1999-08-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-372-044-10

Query Match      100.0%; Score 442; DB 9; Length 156;
Best Local Similarity 100.0%; Pred. NO. 3e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSPSDYALDLSTFLQHPAAF 86
Db 131 PFNLTSPSDYALDLSTFLQHPAAF 156

RESULT 4
US-09-934-249-5
; Sequence 5, Application US/09934249
; Patent No. US20020115081A1
; GENERAL INFORMATION:
; APPLICANT: Lee, Richard T.
; APPLICANT: Landschulz, Katherine T.
; APPLICANT: Turi, Thomas G.
; APPLICANT: Thompson, John P.
; APPLICANT: Kennedy, Scott P.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
; FILE REFERENCE: 20738/7001/ERP/XA
; CURRENT APPLICATION NUMBER: US/09/934,249
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: US 60/227,159
; PRIOR FILING DATE: 2000-08-22
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 156
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-934-249-5

Query Match      100.0%; Score 442; DB 9; Length 156;
Best Local Similarity 100.0%; Pred. NO. 3e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 71 VRRLPVEEPNPAKLLFLILLITVFCQILMAEGVPAPLPEDAPNAASLAPTPVSPVLE 130

QY 61 PFNLTSPSDYALDLSTFLQHPAAF 86

```

Db 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 5

US-09-560-150-11  
; Sequence 11, Application US/09560150  
; Publication No. US20030073076A1  
; GENERAL INFORMATION:  
; APPLICANT: FALB, Dean A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; FILE REFERENCE: 7853-126  
; CURRENT APPLICATION NUMBER: US/09/560,150  
; PRIOR FILING DATE: 2000-04-28  
; PRIOR APPLICATION NUMBER: 09/126,640  
; PRIOR FILING DATE: 1998-07-30  
; PRIOR APPLICATION NUMBER: 08/870,434  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 08/799,910  
; PRIOR FILING DATE: 1997-02-13  
; PRIOR APPLICATION NUMBER: 60/011,787  
; PRIOR FILING DATE: 1996-02-16  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-560-150-11

Query Match 100.0%; Score 442; DB 10; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3e-36; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 0;  
QY 1 VRRQLPVEEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPVLE 60  
Db 71 VRRQLPVEEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPVLE 130  
QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
Db 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 6

US-10-067-741-11  
; Sequence 11, Application US/10067741  
; Publication No. US20030097668A1  
; GENERAL INFORMATION:  
; APPLICANT: Dean A. Falb  
; APPLICANT: Katherine Galvin  
; APPLICANT: Michael Donovan  
; APPLICANT: Dennis Huszar  
; APPLICANT: Michael A. Gimbrome, Jr.  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and  
; Diagnosis of  
; TITLE OF INVENTION: Cardiovascular Disease  
; FILE REFERENCE: 7853-140-999  
; CURRENT APPLICATION NUMBER: US/10/067,741  
; CURRENT FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US/09/288,292  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: 08/870,434  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 08/799,910  
; PRIOR FILING DATE: 1997-02-13  
; PRIOR APPLICATION NUMBER: 60/011,787  
; PRIOR FILING DATE: 1996-02-16  
; PRIOR APPLICATION NUMBER: 08/485,573  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 08/386,844  
; PRIOR FILING DATE: 1995-02-10  
; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 11  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-067-741-11

Query Match 100.0%; Score 442; DB 14; Length 156;  
Best Local Similarity 100.0%; Pred. No. 3e-36; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 0;  
QY 1 VRRQLPVEEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPVLE 60  
Db 71 VRRQLPVEEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPVLE 130  
QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
Db 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 7

US-10-106-698-5647  
; Sequence 5647, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptid  
; FILE REFERENCE: PA005PI  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03  
; NUMBER OF SEQ ID NOS: 8564  
; SOFTWARE: PatentIn Ver. 3.0  
; SEQ ID NO 5647  
; LENGTH: 156  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-106-698-5647

Query Match 100.0%; Score 442; DB 14; Length 166;  
Best Local Similarity 100.0%; Pred. No. 3.2e-36; Indels 0; Gaps 0;  
Matches 86; Conservative 0; Mismatches 0;  
QY 1 VRRQLPVEEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPVLE 60  
Db 81 VRRQLPVEEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPEDAPNAASLAPTPVSPVLE 140  
QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86  
Db 141 PFNLTSEPSDYALDLSTFLQHPAAF 166

RESULT 8

US-10-106-698-6726  
; Sequence 6726, Application US/10106698  
; Publication No. US20030109690A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptid  
; FILE REFERENCE: PA005PI  
; CURRENT APPLICATION NUMBER: US/10/106,698  
; CURRENT FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: PCT/US00/26524  
; PRIOR FILING DATE: 2000-09-28  
; PRIOR APPLICATION NUMBER: US 60/157,137  
; PRIOR FILING DATE: 1999-09-29  
; PRIOR APPLICATION NUMBER: US 60/163,280  
; PRIOR FILING DATE: 1999-11-03

```

; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 6726
; LENGTH: 83
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: (1)_FEATURE
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (7)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (27)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (30)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (43)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (54)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (59)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (60)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (80)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; US-106-698-6726

```

```

Query Match 58.6%; Score 259; DB 14; Length 83;
Best Local Similarity 89.5%; Pred. No. 1.9e-18;
Matches 51; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 30 MAEGVPAPLPEDAPNAASLAPTVSPVLEPFLMTSEPSDYALDLSTFLQQPDAF 86
Db 2 MAEGVPAPLPEDAPNAASLAPTVSPVLEPFLMTSEPSDYALDLSTFLQQPDAF 58

```

```

RESULT 9
US-10-258-106-7
; Sequence 7, Application US/10258106
; Publication No. US20040018185A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; YUE, Henry
; APPLICANT: GANDHI, Amena R.; TRIBOULEY, Catherine M.
; APPLICANT: KEARNEY, Liam; GRIFFIN, Jennifer A.
; APPLICANT: NGUYEN, Daniel B.; BANDMAN, Olga
; APPLICANT: LU, Dying Aina M.; IAL, Preeti G.
; APPLICANT: BURFORD, Neil; KEAN, Farrah A.
; APPLICANT: WALIA, Narinder K.; YAO, Monique G.
; APPLICANT: PATTERSON, Chandra; BURRILL, John D.
; APPLICANT: MARCUS, Gregory A.; ZINGLER, Kurt A.
; APPLICANT: RECIPON, Shixley A.; LU, Yan
; APPLICANT: POLICKY, Jennifer L.; THORNTON, Michael B.
; APPLICANT: TANG, Y. Tom; HAFALIA, April J.A.
; APPLICANT: ELLIOTT, Vicki S.; BAUGHN, Mariah R.
; APPLICANT: WALSH, Roderick T.; RAMKUMAR, Jayalaxmi
; APPLICANT: PCROWSKY, Mark L.; AU-YOUNG, Janice K.
; APPLICANT: HILLMAN, Jennifer L.; GURURAJAN, Rajagopal
; TITLE OF INVENTION: HUMAN KINASES
; FILE REFERENCE: PI-0076 USN
; CURRENT APPLICATION NUMBER: US/10/258,106
; CURRENT FILING DATE: 2002-10-18

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/12992
; PRIOR FILING DATE: 2001-04-20
; PRIOR APPLICATION NUMBER: US 60/199,021
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/200,226
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/202,339
; PRIOR FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/203,505
; PRIOR FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: US 60/205,654
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 60/207,739
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 60/208,795
; PRIOR FILING DATE: 2000-06-01
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PERL Program
; SEQ ID NO 7
; LENGTH: 923
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040018185A1 2589355CD1
; US-10-258-106-7

```

```

Query Match 19.1%; Score 84.5; DB 15; Length 923;
Best Local Similarity 33.3%; Pred. No. 5.5;
Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;

```

```

QY 7 VEENPAKELF-----LIVFCQILMAEGVP-----APLPEDAPNAASLAP----- 52
Db 709 LQHERLQQKFLQKQSQLOQHFNQIAESSYQPSQQLPLPRQETPPSQQAPPSLT 768
QY 53 TPVSPVLEPFLMTSEPSDYALDLSTFLQQH 82
Db 769 QPLSPVLEP---SSEQMQY---SPFLSQY 791

```

```

RESULT 10
US-10-051-874-166
; Sequence 166, Application US/10051874
; Publication No. US2004000557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Alsobrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Boldog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: Macbougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glenda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Herrman, John L
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh

```

```
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
; APPLICANT: Burgess, Catherine E
; TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF
; FILE OF INVENTION: USING THE SAME
; FILE REFERENCE: 21402-245
; CURRENT APPLICATION NUMBER: US/10/051,874
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: 60/268,595
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/325,306
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/262,587
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/272,409
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: 60/262,454
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: 60/276,777
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/291,672
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 60/330,336
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/265,530
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/261,376
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 269
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 166
; LENGTH: 5008
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-051-874-156

Query Match 17.6%; Score 78; DB 15; Length 5008;
Best Local Similarity 30.4%; Pred. No. 1.6e+02;
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRQLPVEPNPAKRLFLLLTIVFCQIIAEEGVPAPLPPEDAPNAASLAPTVSPVLE 60
D5 4350 LKQESAPEPTQHYTYNVSNLDVRL-----SAPPPPEPSPPLAPSPAPPTTE 4402
QY 61 PF-NLTSEP 68
D5 4403 PLVELTEP 4411

RESULT 11
US-10-085-198-112
; Sequence 112, Application US/10085198
; Publication No. US20040009907A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook et al.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-279
; CURRENT APPLICATION NUMBER: US/10/085,198
; CURRENT FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/271,646
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 60/276,401
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/311,981
; PRIOR FILING DATE: 2001-08-13
; PRIOR APPLICATION NUMBER: 60/312,858
; PRIOR FILING DATE: 2001-08-16
; PRIOR APPLICATION NUMBER: 60/271,840
```

```
; PRIOR FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 60/277,324
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 60/286,096
; PRIOR FILING DATE: 2001-04-21
; PRIOR APPLICATION NUMBER: 60/299,695
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/315,614
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/272,405
; PRIOR FILING DATE: 2001-02-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 653
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 112
; LENGTH: 5159
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-085-198-112

Query Match 17.6%; Score 78; DB 15; Length 5159;
Best Local Similarity 30.4%; Pred. No. 1.7e+02;
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRQLPVEPNPAKRLFLLLTIVFCQIIAEEGVPAPLPPEDAPNAASLAPTVSPVLE 60
D5 4299 LKQESAPEPTQHYTYNVSNLDVRL-----SAPPPPEPSPPLAPSPAPPTTE 4351
QY 61 PF-NLTSEP 68
D5 4352 PLVELTEP 4360

RESULT 12
US-10-051-874-165
; Sequence 165, Application US/10051874
; Publication No. US2004000557A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Alsbrook II, John P
; APPLICANT: Colman, Steven D
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Bollog, Ferenc
; APPLICANT: Vernet, Corine AM
; APPLICANT: Li, Li
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Casman, Stacie J
; APPLICANT: Guo, Xiaojia Sasha
; APPLICANT: Edinger, Shlomit R
; APPLICANT: MacDougall, John R
; APPLICANT: Malyankar, Uriel M
; APPLICANT: Patturajan, Meera
; APPLICANT: Shinkets, Richard A
; APPLICANT: Pena, Carol EA
; APPLICANT: Tchernev, Vellizar T
; APPLICANT: Zerhusen, Bryan D
; APPLICANT: Millet, Isabelle
; APPLICANT: Miller, Charles E
; APPLICANT: Lepley, Denise M
; APPLICANT: Smithson, Glennda
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Heriman, John I
; APPLICANT: Peyman, John A
; APPLICANT: Gorman, Linda
; APPLICANT: Mezes, Peter D
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Gerlach, Valerie
; APPLICANT: Grosse, William M
; APPLICANT: Liu, Xiaohong
; APPLICANT: Ellerman, Karen
; APPLICANT: Rothenberg, Mark
; APPLICANT: Stone, David J
```

APPLICANT: Burgess, Catherine E  
TITLE OF INVENTION: PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF  
TITLE OF INVENTION: USING THE SAME  
FILE REFERENCE: 21402-245  
CURRENT FILING DATE: 2002-09-25  
PCT APPLICATION NUMBER: US/10/051,874  
CURRENT FILING DATE: 2002-09-25  
PCT APPLICATION NUMBER: 60/268,595  
PCT FILING DATE: 2001-02-14  
PCT APPLICATION NUMBER: 60/325,306  
PCT FILING DATE: 2001-09-27  
PCT APPLICATION NUMBER: 60/262,587  
PCT FILING DATE: 2001-01-18  
PCT APPLICATION NUMBER: 60/272,409  
PCT FILING DATE: 2001-02-28  
PCT APPLICATION NUMBER: 60/262,454  
PCT FILING DATE: 2001-01-18  
PCT APPLICATION NUMBER: 60/276,777  
PCT FILING DATE: 2001-03-16  
PCT APPLICATION NUMBER: 60/291,672  
PCT FILING DATE: 2001-05-17  
PCT APPLICATION NUMBER: 60/330,336  
PCT FILING DATE: 2001-10-18  
PCT APPLICATION NUMBER: 60/265,530  
PCT FILING DATE: 2001-01-31  
PCT APPLICATION NUMBER: 60/261,376  
PCT FILING DATE: 2001-01-16  
NUMBER OF SEQ ID NOS: 269  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 165  
LENGTH: 5262  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-051-874-165

Query Match 17.68; Score 78; DB 15; Length 5262;  
Best Local Similarity 30.44; Pred. No. 1.7e+02;  
Matches 21; Conservative 19; Mismatches 28; Indels 8; Gaps 2;

Qy 1 VRQLPVEPNPAKRLIFLLITVFCQILMAEGVPAPLPPEDAPNAASLAPTPVSPVL 60  
Db 4604 LKQESAPPTQHQYIYVNSLDVRL-----SAPPRESPPPSFLAPSPAPPTPE 4656

Qy 61 PF-NLTSEP 68  
Db 4657 PLVELPTER 4665

RESULT 13  
US-09-864-761-36180  
Sequence 36180, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Acomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-09-23  
PCT APPLICATION NUMBER: US 60/180,312  
PCT FILING DATE: 2000-02-04  
PCT APPLICATION NUMBER: US 60/207,456  
PCT FILING DATE: 2000-05-26  
PCT APPLICATION NUMBER: US 09/632,366  
PCT FILING DATE: 2000-08-03  
PCT APPLICATION NUMBER: GB 24263.6  
PCT FILING DATE: 2000-10-04  
PCT APPLICATION NUMBER: US 60/236,359  
PCT FILING DATE: 2000-09-27  
PCT APPLICATION NUMBER: PCT/US01/00666  
PCT FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/00667  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00664  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00669  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00665  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00668  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00663  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00662  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00661  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: PCT/US01/00670  
PCT FILING DATE: 2001-01-30  
PCT APPLICATION NUMBER: US 60/234,687  
PCT FILING DATE: 2000-09-21  
PCT APPLICATION NUMBER: US 09/608,408  
PCT FILING DATE: 2000-06-30  
PCT APPLICATION NUMBER: US 09/774,203  
PCT FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 36180  
LENGTH: 242  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL022333.1  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.5  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9  
US-09-864-761-36180

Query Match 17.1%; Score 75.5; DB 9; Length 242;  
Best Local Similarity 35.7%; Pred. No. 9.4; Indels 3; Gaps 2;  
Matches 20; Conservative 10; Mismatches 23

Qy 5 LPVEPNPAKRLIFLLITVFCQILMAEGVPAPL-PPEDAPNAASLAPTPVSPVL 59  
Db 41 LPVPPSSSTSSLLFLALLFLPLFL-LLPPLFPPLPPSFLPPPLSLPL 94

RESULT 14  
US-09-298-523B-8  
Sequence 8, Application US/09298523B  
Publication No. US20030059438A1  
GENERAL INFORMATION:  
APPLICANT: BRILES et al.  
TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS  
FILE REFERENCE: 454312-3140  
CURRENT APPLICATION NUMBER: US/09/298,523B  
CURRENT FILING DATE: 1999-04-23  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 8  
LENGTH: 434  
TYPE: PRT  
ORGANISM: Streptococcus pneumoniae  
US-09-298-523B-8

Query Match 17.1%; Score 75.5; DB 10; Length 434;

```
Best Local Similarity 24.5%; Pred. No. 18;
Matches 24; Conservative 10; Mismatches 25; Indels 39; Gaps 4;

QY 3 ROLPVEFNPAKBLLELLTIVFCQILMAEAGVPALP--PEDAPNAASLAPTPVSPVLE 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 345 KEKPAEQQPA-----PAPQEKTPERPENPAPAKPENPAPK 382

QY 61 PNLINSEP-----SDYAL-----DLSTFLOQHP 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 PENPAPQPKAEKADQQAEDYARRSEFYNRLTQQOP 420

RESULT 15
US-09-864-761-39500
; Sequence 39500, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39500
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005483.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.8
```

```
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.2
US-09-864-761-39500

Query Match 16.4%; Score 72.5; DB 9; Length 111;
Best Local Similarity 36.7%; Pred. No. 7.7;
Matches 22; Conservative 5; Mismatches 22; Indels 11; Gaps 2;

QY 26 CQILMAEAGVPALPPEDAPNAASLAPTPVSPVLEPNTSEPSYALDLSTFLOQHPAA 85
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 CSYLIPD--LEAPAPAPAPAAAAAASPVNPAPATSPTCALAV-----PAA 53

Search completed: February 25, 2004, 05:41:39
Job time : 35 secs
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model  
Run on: February 25, 2004, 05:32:57 ; Search time 21 Seconds  
(without alignments)  
393.927 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156  
Perfect score: 442  
Sequence: 1 VRRQLPVEPNPAKRLFL...EPDYALDLSTFLQHPAPF 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues  
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 78:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | ID       | Description        |
|------------|-------|---------------|--------|----------|--------------------|
| 1          | 442   | 100.0         | 156    | 2 JCS537 | differentiation-de |
| 2          | 273   | 61.8          | 153    | 2 S33363 | gly96 protein - mo |
| 3          | 81.5  | 18.4          | 102    | 2 T33565 | hypothetical prote |
| 4          | 79.5  | 18.0          | 1596   | 2 A41216 | guanine nucleotide |
| 5          | 78    | 17.6          | 4957   | 2 T03455 | ALR protein - huma |
| 6          | 78    | 17.6          | 5262   | 2 T03454 | ALR protein - huma |
| 7          | 76    | 17.2          | 148    | 2 A81141 | acyl CoA thioester |
| 8          | 76    | 17.2          | 1677   | 2 T18344 | P-glycoprotein B - |
| 9          | 74    | 16.7          | 955    | 2 T00247 | zinc finger protei |
| 10         | 74    | 16.7          | 1561   | 2 T00248 | zinc finger protei |
| 11         | 73.5  | 16.6          | 272    | 2 A47021 | pectic enzyme secr |
| 12         | 73.5  | 16.6          | 925    | 2 S27920 | nuclear antigen EB |
| 13         | 73    | 16.5          | 1029   | 2 T30351 | mucin-like protein |
| 14         | 72.5  | 16.4          | 272    | 2 S28013 | outC protein - Erw |
| 15         | 71.5  | 16.2          | 626    | 1 NBHUIA | platelet glycoprot |
| 16         | 71    | 16.1          | 266    | 2 B87498 | hypothetical prote |
| 17         | 71    | 16.1          | 359    | 2 T33756 | hypothetical prote |
| 18         | 70.5  | 16.0          | 235    | 2 D83477 | hypothetical prote |
| 19         | 70.5  | 16.0          | 505    | 1 Q0B529 | BRF1 protein - hu  |
| 20         | 70    | 15.8          | 148    | 2 X81878 | probable acyl-CoA  |
| 21         | 70    | 15.8          | 1233   | 2 T37045 | nitrate reductase  |
| 22         | 69.5  | 15.7          | 1887   | 2 S61703 | fatty-acid synthas |
| 23         | 69    | 15.6          | 328    | 2 AB0810 | cell division prot |
| 24         | 69    | 15.6          | 418    | 2 T19800 | hypothetical prote |
| 25         | 69    | 15.6          | 1212   | 2 D88175 | protein T24H7.5a [ |
| 26         | 69    | 15.6          | 1402   | 2 T46707 | translation initia |
| 27         | 69    | 15.6          | 1454   | 2 C88175 | protein T24H7.5b [ |
| 28         | 69    | 15.6          | 2187   | 2 T30826 | nascent polypeptid |
| 29         | 68.5  | 15.5          | 181    | 2 T08793 | hypothetical prote |

|    |      |      |      |          |                     |
|----|------|------|------|----------|---------------------|
| 30 | 68.5 | 15.5 | 882  | 1 IJHUCE | cadherin 1 precurs  |
| 31 | 68.5 | 15.5 | 1072 | 2 T50949 | verprolin related   |
| 32 | 68.5 | 15.5 | 1337 | 1 I38670 | protein-tyrosine-p  |
| 33 | 68.5 | 15.5 | 1844 | 2 S01956 | hypothetical prote  |
| 34 | 68.5 | 15.5 | 2649 | 2 T51023 | hypothetical prote  |
| 35 | 68   | 15.4 | 211  | 2 B89716 | protein F45B8.3 [i  |
| 36 | 68   | 15.4 | 241  | 2 T22216 | hypothetical prote  |
| 37 | 68   | 15.4 | 267  | 2 A49068 | craniosynostosis-as |
| 38 | 68   | 15.4 | 1396 | 2 A44453 | translation initia  |
| 39 | 67.5 | 15.3 | 431  | 2 D87518 | lysosyme family pr  |
| 40 | 67   | 15.2 | 142  | 2 E84369 | histidine triad pr  |
| 41 | 67   | 15.2 | 283  | 2 S13383 | hydroxyproline-ric  |
| 42 | 67   | 15.2 | 586  | 2 S19381 | hypothetical prote  |
| 43 | 67   | 15.2 | 3566 | 1 A40701 | tenascin-X precurs  |
| 44 | 66.5 | 15.0 | 169  | 2 A72466 | hypothetical prote  |
| 45 | 66.5 | 15.0 | 3149 | 1 Q0B58  | BRF1 protein - hu   |

ALIGNMENTS

RESULT 1

JCS537  
differentiation-dependent protein DIF-2 - human  
C;Species: Homo sapiens (man)  
C;Date: 02-Sep-1997 #sequence\_revision 05-Sep-1997 #text\_change 07-May-1999  
C;Accession: JCS537  
R;Pietzsch, A.; Buechler, C.; Aslanidis, C.; Schmitz, G.  
Biochem. Biophys. Res. Commun. 235, 4-9, 1997  
A;Title: Identification and characterization of a novel monocyte/macrophage differentia  
A;Reference number: JCS537; MUID:97339426; PMID:9196025  
A;Accession: JCS537  
A;Molecule type: mRNA  
A;Residues: 1-156 <PIE>  
A;Experimental source: monocyte  
A;Note: the authors translated the codon CCG for residue 106 as Arg  
Query Match 100.0%; Score 442; DB 2; Length 156;  
Best Local Similarity 100.0%; Pred. No. 7.1e-36;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |     |   |     |
|----|-----|---|-----|
| Qy | 1   | VRRQLPVEPNPAKRLFLLLTLVFCQILMAEGVPAPLPDPEDAPNAASLAPTPSPVLE | 60  |
| Db | 71  | VRRQLPVEPNPAKRLFLLLTLVFCQILMAEGVPAPLPDPEDAPNAASLAPTPSPVLE | 130 |
| Qy | 61  | PNLTSPSDYALDLSTFLQHPAPF                                   | 86  |
| Db | 131 | PNLTSPSDYALDLSTFLQHPAPF                                   | 156 |

RESULT 2

S33363  
gly36 protein - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 16-Feb-1997  
C;Accession: S33363  
R;Charles, C.H.; Yoon, J.K.; Sinske, J.S.; Lau, L.F.  
Oncogene 8, 797-801, 1993  
A;Title: Genomic structure, cDNA sequence, and expression of gly96, a growth factor-inc  
A;Reference number: S33363; MUID:93173526; PMID:8437864  
A;Accession: S33363  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-153 <CHA>  
A;Cross-references: EMBL:X67644  
C;Genetics:  
A;Introns: 70/3  
C;Keywords: transmembrane protein  
Query Match 61.8%; Score 273; DB 2; Length 153;  
Best Local Similarity 68.2%; Pred. No. 1.6e-19;  
Matches 58; Conservative 5; Mismatches 14; Indels 8; Gaps 2;



QY 1 VRRQLPVEENPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTVS----- 56  
 DB 71 VRRQLPVEENPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTVS----- 128  
 QY 57 --PVLEPNLTSPSDYALDLSTEL 79  
 DB 129 APPVLEPNLTSPSDYALDLKAF 153

RESULT 3  
 T33565  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 18-Feb-2000  
 C:Accession: T33565  
 R:Nelson, J.; Gattung, S.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of C. elegans cosmid R160.  
 A:Reference number: Z1370  
 A:Accession: T33565  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-102 <NLS>  
 A:Cross-references: EMBL:AF099001; PIDN:AAC68734.1; GSPDB:GN00028; CESP:R160.5  
 A:Experimental source: strain Bristol N2; clone R160  
 C:Genetics:  
 A:Gene: CESP:R160.5  
 A:Map position: X  
 A:Introns: 25/2

Query Match 18.4%; Score 81.5; DB 2; Length 102;  
 Best Local Similarity 24.7%; Pred. No. 0.37; Mismatches 11; Gaps 3;  
 Matches 19; Conservative 19;  
 QY 11 NPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTVSPVLEPN-- 63  
 DB 7 NPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTVSPVLEPN-- 63  
 QY 64 ---LTSEPSDYALDLST 77  
 DB 66 QWKIDPTESDVDMTS 82

RESULT 4  
 A41216  
 C:Species: Drosophila melanogaster  
 C:Date: 28-May-1992 #sequence\_revision 28-May-1992 #text\_change 24-Sep-1998  
 C:Accession: A41216  
 R:Simon, M.A.; Bowtell, D.D.L.; Dodson, G.S.; Lavery, T.R.; Rubin, G.M.  
 Cell 67, 701-716, 1991  
 A:Title: Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in  
 A:Reference number: A41216; MUID:92034991; PMID:1934068  
 A:Accession: A41216  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1596 <Sim>  
 A:Cross-references: GB:M77501; MID:g158470; PID:g158471  
 C:Genetics:  
 A:Gene: FlyBase:Scs  
 A:Cross-references: FlyBase:Scs  
 C:Superfamily: CDC25-type guanine nucleotide exchange activator homology; pleckstrin rep  
 F:1479-586/Domain: pleckstrin repeat homology <PLK>  
 F:825-1066/Domain: CDC25-type guanine nucleotide exchange activator homology <SCS>

Query Match 18.0%; Score 79.5; DB 2; Length 1596;  
 Best Local Similarity 26.3%; Pred. No. 11;  
 Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;  
 QY 1 VRRQLPVE-----EPNPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTV 53  
 DB 1423 IRRNSAIEKRAAATSPNQAAAGPSTTLTVNSQAVATDEPLPLPISP-----AASGTT 1477

QY 54 --PVSPVLEP--NLTSEP-----SDYALDLSTFLQ-----HPAAF 86  
 DB 1478 TSLTPAMSPNIPSHPVSTSSVAHQLRWQCOQQOQTHPAIY 1523  
 RESULT 5  
 T03455  
 ALR protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 27-Oct-2003  
 C:Accession: T03455  
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano  
 Oncogene 15, 549-560, 1997  
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolo  
 A:Reference number: Z14954; MUID:97388474; PMID:9247308  
 A:Accession: T03455  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-4957 <PRA>  
 A:Cross-references: EMBL:AF010404; NID:g2358286; PIDN:AAC51735.1; PID:g2358287  
 C:Genetics:  
 A:Gene: ALR  
 A:Map position: 12  
 C:Superfamily: acute lymphoblastic leukemia protein, ALR type  
 C:Keywords: alternative splicing

Query Match 17.6%; Score 78; DB 2; Length 4957;  
 Best Local Similarity 30.4%; Pred. No. 55;  
 Matches 21; Conservative 21; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRRQLPVEENPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTVSPVLE 60  
 DB 4299 LKQSPAPPEPTQHRVTYVNSLDVQL-----SAPPEPSPPPSLAPSPSPTE 4351  
 QY 61 PF-NLTSEP 68  
 DB 4352 FLVELPTEP 4360

RESULT 6  
 T03454  
 ALR protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 27-Oct-2003  
 C:Accession: T03454  
 R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yanc  
 Oncogene 15, 549-560, 1997  
 A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolo  
 A:Reference number: Z14954; MUID:97388474; PMID:9247308  
 A:Accession: T03454  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-5262 <PRA>  
 A:Cross-references: EMBL:AF010403; NID:g2358284; PIDN:AAC51734.1; PID:g2358285  
 C:Genetics:  
 A:Gene: ALR  
 A:Map position: 12  
 C:Superfamily: acute lymphoblastic leukemia protein, ALR type  
 C:Keywords: alternative splicing

Query Match 17.6%; Score 78; DB 2; Length 5262;  
 Best Local Similarity 30.4%; Pred. No. 58;  
 Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRRQLPVEENPAKRLIFLLITVFCQILMAEGVPAPLPDPADPAASLAPTVSPVLE 60  
 DB 4604 LKQSPAPPEPTQHRVTYVNSLDVQL-----SAPPEPSPPPSLAPSPSPTE 4656  
 QY 61 PF-NLTSEP 68  
 DB 4657 FLVELPTEP 4665

RESULT 7  
A81141  
acyl CoA thioester hydrolase family protein NME0925 [imported] - Neisseria meningitidis  
C:Species: Neisseria meningitidis  
C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: A81141  
R:Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.  
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;  
Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000  
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve  
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755; PMID:10710307  
A:Accession: A81141  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-148 <TEF>  
A:Cross-references: GB:AE002444; GB:AE002098; NID:g7226162; PIDN:AAF41333.1; PID:g722616  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NME0925

Query Match 17.2%; Score 76; DB 2; Length 148;  
Best Local Similarity 40.5%; Pred. No. 1.9;  
Matches 15; Conservative 5; Mismatches 17; Indels 0; Gaps 0;  
QY 18 FLLIIVFCQLMAEGVAPLPEDAPNAASLPTP 54  
DB 112 YIVTAVFYVAIDAGNRPPIKGNPKXLAGLPTP 148

RESULT 8  
T18344  
P-Glycoprotein B - Leishmania tropica  
C:Species: Leishmania tropica  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T18344  
R:La Fuente, E.; Castany, S.; Gamarro, F.  
submitted to the EMBL Data Library, April 1996  
A:Reference number: Z18880  
A:Accession: T18344  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1677 <LAF>  
A:Cross-references: EMBL:U55381; NID:g4916605; PID:g4916606; PIDN:AAB51191.1  
C:Superfamily: human multidrug resistance protein cMOAT; ATP-binding cassette homology

Query Match 17.2%; Score 76; DB 2; Length 1677;  
Best Local Similarity 33.9%; Pred. No. 26;  
Matches 20; Conservative 6; Mismatches 15; Indels 18; Gaps 2;

QY 37 APDP-----PEDAPNAASLPTPVSPVLEPN-----LTSEPSDYALDLS 77  
DB 698 APLPEAEAPLPDPFSSSAAPAGPVTEPLNKGSHGHDAAASSEPLSSSAQKST 756

RESULT 9  
T00247  
zinc finger protein wiz - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
C:Accession: T00247  
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.  
submitted to the EMBL Data Library, March 1998  
A:Description: Molecular cloning and distinct developmental expression pattern of splice  
A:Reference number: Z14130  
A:Accession: T00247  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-955 <MAT>  
A:Cross-references: EMBL:AB012266; NID:dl227741; PIDN:BAA32791.1; PID:dl033757  
A:Experimental source: brain  
C:Genetics:

A:Gene: wiz

Query Match 16.7%; Score 74; DB 2; Length 955;  
Best Local Similarity 41.9%; Pred. No. 22;  
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;  
QY 34 GVPAPLPEDAPNAASLPTPVSPVLE-----PFNLTSEP 68  
DB 300 GSTPKNPKDKSPQLSLSRPSPRAQCPQSEDEGLNLTSGP 342

RESULT 10  
T00248  
zinc finger protein wizL - mouse  
C:Species: Mus musculus (house mouse)  
C>Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 05-Nov-1999  
C:Accession: T00248  
R:Matsumoto, K.; Ishii, N.; Yoshida, S.; Shiosaka, S.; Wanaka, A.; Tohyama, M.  
submitted to the EMBL Data Library, March 1998  
A:Description: Molecular cloning and distinct developmental expression pattern of splice  
A:Reference number: Z14130  
A:Accession: T00248  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1561 <MAT>  
A:Cross-references: EMBL:AB012265; NID:dl227740; PIDN:BAA32790.1; PID:dl033756  
A:Experimental source: brain  
C:Genetics:  
A:Gene: wiz

Query Match 16.7%; Score 74; DB 2; Length 1561;  
Best Local Similarity 41.9%; Pred. No. 38;  
Matches 18; Conservative 2; Mismatches 15; Indels 8; Gaps 1;

QY 34 GVPAPLPEDAPNAASLPTPVSPVLE-----PFNLTSEP 68  
DB 906 GSTPKNPKDKSPQLSLSRPSPRAQCPQSEDEGLNLTSGP 948

RESULT 11

A47021  
pectic enzyme secretion protein OutC - Erwinia chrysanthemi  
C:Species: Erwinia chrysanthemi  
C>Date: 24-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 16-Feb-1997  
C:Accession: A47021  
R:Lindeberg, M.; Collmer, A.  
J. Bacteriol. 174, 7385-7397, 1992  
A:Title: Analysis of eight out genes in a cluster required for pectic enzyme secretion  
A:Reference number: A47021; MUID:93054355; PMID:1429461  
A:Accession: A47021  
A>Status: preliminary; not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-272 <LIN>  
A:Note: sequence extracted from NCBI backbone (NCBIP:118270)  
C:Keywords: transmembrane protein

Query Match 16.6%; Score 73.5; DB 2; Length 272;  
Best Local Similarity 30.1%; Pred. No. 6.4;  
Matches 22; Conservative 13; Mismatches 25; Indels 13; Gaps 3;

QY 4 QLPVEENPAKLLPILLITVFCQ----ILMAEGVPAPLPEDAPNAASLPTPVSPVLE 60  
DB 5 KLPPSPSVTRILPILYLLMLLFCQLAMIFWRVGLP-----DNSPVSIVQTTPAARQQ 59

QY 61 PFNLTSEPSDYAL 73  
DB 59 PVTL----NDFTL 67

RESULT 12

S27920  
nuclear antigen EBNA-3A - human herpesvirus 4  
C:Species: human herpesvirus 4, Epstein-Barr virus

C>Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text\_change 21-Jul-2000  
 C/Accession: S27920; G49253; H49253; F49253  
 R/Sample: J.; Young, L.; Martin, B.; Chatman, T.; Kieff, E.; Kieff, E.  
 Submitted to the EMBL Data Library, July 1990  
 A/Reference number: S27920  
 A/Accession: S27920  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-925 <SAM>  
 A/Cross-references: EMBL:W34440; NID:G330407; PIDN:AAA45893.1; PID:G330408  
 R/Apolon, A.; Moss, D.; Stumm, R.; Burrows, S.; Suhrbier, A.; Misko, I.; Schmidt, C.;  
 Eur. J. Immunol. 22, 183-189, 1992  
 A/Title: Sequence variation of cytotoxic T cell epitopes in different isolates of Epstein  
 A/Reference number: A49034; MUID:92111623; PMID:1370413  
 A/Accession: G49253  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 314-332 <APO>  
 A/Cross-references: GB:S79235; NID:G242901; PIDN:AA20990.1; PID:G242902  
 A/Experimental source: human B-type strain QIMR-JSM6  
 A/Note: sequence extracted from NCBI backbone (NCBIN:79235, NCBIIP:79252)  
 A/Accession: H49253  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 314-338 <AP2>  
 A/Cross-references: GB:S79234; NID:G242899; PIDN:AA20989.1; PID:G242900  
 A/Experimental source: human B-type strain AG876US; human B-type strain L4WR  
 A/Note: sequence extracted from NCBI backbone (NCBIN:79237, NCBIIP:79254, NCBIN:79234, NC  
 C/Genetics: 112/3  
 A/Intros: 112/3  
 A/Superfamily: human herpesvirus 4 nuclear antigen EBNA-3A

Query Match 16.6%; Score 73.5; DB 2; Length 925;  
 Best Local Similarity 28.4%; Pred. NC. 24;  
 Matches 29; Conservative 13; Mismatches 39; Indels 21; Gaps 5;  
 QY 1 VRQPLPVEEPNPKRLFLILLITVFCQI--LMAEGVPAPLP-----PEDA 44  
 DB 606 VSPQPMERPEPQQNF--PGSPFSQADVARESGVPANQVFDPLTPQISGAPAA 663  
 QY 45 ENAAASLATPVPSPVLRP--FNL--TSEPSDYALDLSTFLQHP 83  
 DB 664 FLRASGVPVPVPAQVQVFDPLTEPINCASNAHFLQQP 705

RESULT 13  
 T30351  
 C/Species: Lymantria dispar nuclear polyhedrosis virus  
 C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 R/Kuzio, J.; Pearson, M.N.; Harwood, S.H.; Funk, C.J.; Evans, J.T.; Slavicek, J.M.; Rohr  
 Virol. 253, 17-34, 1999  
 A/Title: Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria d  
 A/Reference number: Z20836; MUID:99124785; PMID:9887315  
 A/Accession: T30351  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1029 <KUZ>  
 A/Cross-references: EMBL:AF081810; PIDN:AA070189.1

Query Match 16.5%; Score 73; DB 2; Length 1029;  
 Best Local Similarity 38.2%; Pred. NC. 30;  
 Matches 29; Conservative 3; Mismatches 30; Indels 14; Gaps 4;  
 QY 6 PVPEPNPKRLFLILLITVFCQILMAEGVPAPLP--PEDAPNAASL-----APTPVSP 57  
 DB 644 PVPEPS-APVDVFTLNSAEF-----APEPAPEAPAPAPAPAPAPAPAPAPAP 697  
 QY 58 VLEPFLNLTSEPSDYAL 73  
 DB 698 ALERFVPALEPVEPAL 713

RESULT 14

S28013  
 outC protein - Erwinia chrysanthemi  
 C/Species: Erwinia chrysanthemi  
 C/Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 21-Jul-2000  
 C/Accession: S28013; S23865  
 R/Condemine, G.; Doral, C.; Hugouvieux-Cotte-Pattat, N.; Robert-Baudouy, J.  
 Mol. Microbiol. 6, 3199-3211, 1992  
 A/Title: Some of the cut genes involved in the secretion of pectate lyases in Erwinia ci  
 A/Reference number: S28011; MUID:93086427; PMID:1453958  
 A/Accession: S28013  
 A/Molecule type: DNA  
 A/Residues: 1-272 <CON>  
 A/Cross-references: EMBL:X65265; NID:G3152953; PIDN:CAA46369.1; PID:G42201  
 C/Genetics:  
 A/Gene: outC  
 C/Keywords: transmembrane protein

Query Match 16.4%; Score 72.5; DB 2; Length 272;  
 Best Local Similarity 30.1%; Pred. No. 8;  
 Matches 22; Conservative 16; Mismatches 22; Indels 13; Gaps 4;  
 QY 4 QLPVEEPNPKRLFLILLITVFCQI--LMAEGVPAPLPEDAPNAASLATPVPVLE 60  
 DB 5 KLPPLSPSVIRILFYLLMLFCCQLAMIFWIGL-----PDNAP-VSSVQITPAQAQQ 58  
 QY 61 PNLITSEPSDYAL 73  
 DB 59 PVTL---NDFTL 67

RESULT 15

NBHUIA  
 platelet glycoprotein Ib alpha chain precursor - human  
 N/Alternate names: membrane glycoprotein Ib alpha chain  
 N/Contains: glycosialicin  
 C/Species: Homo sapiens (man)  
 C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 22-Jun-1999  
 C/Accession: A94174; A60435; A94173; S16945; I55355; A27075; A27102  
 R/Lopez, J.A.; Chung, D.W.; Fujikawa, K.; Hagen, F.S.; Papayannopoulou, T.; Roth, G.J.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 5615-5619, 1987  
 A/Title: Cloning of the alpha-chain of human platelet glycoprotein Ib: a transmembrane  
 A/Reference number: A94174; MUID:87289655; PMID:3303030  
 A/Accession: A94174  
 A/Molecule type: mRNA  
 A/Residues: 1-626 <LOP>  
 A/Cross-references: GB:J02940; NID:gl83499; PIDN:AAA52595.1; PID:G306793  
 R/Wicki, A.N.; Walz, A.; Gexber-Huber, S.N.; Wenger, R.H.; Vornhagen, R.; Clemetson, K.  
 Thromb. Haemost. 61, 448-453, 1989  
 A/Title: Isolation and characterization of human blood platelet mRNA and construction o  
 d cloning of a GPIb coding cDNA insert.  
 A/Reference number: A60435; MUID:90020160; PMID:2799758  
 A/Accession: A60435  
 A/Molecule type: mRNA  
 A/Residues: 207-467 <WIC>  
 R/Titani, K.; Takio, K.; Harda, M.; Ruggeri, Z.M.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 5610-5614, 1987  
 A/Title: Amino acid sequence of the von Willebrand factor-binding domain of platelet me  
 A/Reference number: A94173; MUID:87289654; PMID:3497398  
 A/Accession: A94173  
 A/Molecule type: protein  
 A/Residues: 17-315 <TIT>  
 R/Hess, D.; Schaller, J.; Rickli, E.E.; Clemetson, K.J.  
 Eur. J. Biochem. 199, 389-393, 1991  
 A/Title: Identification of the disulphide bonds in human platelet glycosialicin.  
 A/Reference number: S16945; MUID:91301149; PMID:2070794  
 A/Accession: S16945  
 A/Status: preliminary  
 A/Molecule type: protein  
 A/Residues: 224-227; 262-270; 277-282 <HES>  
 R/Lopez, J.A.; Ludwig, E.H.; McCarthy, B.J.  
 J. Biol. Chem. 267, 10055-10061, 1992



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:28:41 ; Search time 14 Seconds  
(without alignments)  
319.860 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442  
Sequence: 1 VRQLPVEEPNPKRLFL.....EPDYALDLSTFLQHPAP 86

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1          | 442   | 100.0       | 156    | 1     | IEXL1 HUMAN |
| 2          | 442   | 100.0       | 156    | 1     | IEXL1 PANTR |
| 3          | 273   | 61.8        | 153    | 1     | IEXL1 MOUSE |
| 4          | 78.5  | 17.8        | 1595   | 1     | SOS DROME   |
| 5          | 78    | 17.6        | 5262   | 1     | MLL2 HUMAN  |
| 6          | 73.5  | 16.6        | 272    | 1     | GSQC ERWCH  |
| 7          | 72.5  | 16.4        | 272    | 1     | GSQC ERWCH  |
| 8          | 72    | 16.3        | 852    | 1     | WS14 HUMAN  |
| 9          | 71.5  | 16.2        | 626    | 1     | GPBA HUMAN  |
| 10         | 70.5  | 16.0        | 605    | 1     | BRL1 EBV    |
| 11         | 70.5  | 16.0        | 2715   | 1     | MLL4 HUMAN  |
| 12         | 69.5  | 15.7        | 914    | 1     | GNDS HUMAN  |
| 13         | 69.5  | 15.7        | 1887   | 1     | FAS2 YEAST  |
| 14         | 69    | 15.6        | 328    | 1     | ZIPA SALTY  |
| 15         | 69    | 15.6        | 646    | 1     | NA95 HUMAN  |
| 16         | 69    | 15.6        | 1402   | 1     | IF4G RABIT  |
| 17         | 68.5  | 15.5        | 181    | 1     | CU25 HUMAN  |
| 18         | 68.5  | 15.5        | 882    | 1     | CADI1 HUMAN |
| 19         | 68.5  | 15.5        | 1337   | 1     | PTPJ HUMAN  |
| 20         | 68.5  | 15.5        | 1844   | 1     | POLR1 TYMV  |
| 21         | 68    | 15.4        | 267    | 1     | MSX2 HUMAN  |
| 22         | 68    | 15.4        | 1395   | 1     | IF4G HUMAN  |
| 23         | 67.5  | 15.3        | 426    | 1     | ASD4 NEUCR  |
| 24         | 67    | 15.2        | 283    | 1     | EXTN1 SORBI |
| 25         | 67    | 15.2        | 586    | 1     | LRE1 YEAST  |
| 26         | 67    | 15.2        | 802    | 1     | ENAH MOUSE  |
| 27         | 67    | 15.2        | 4289   | 1     | TENX HUMAN  |
| 28         | 66.5  | 15.0        | 1505   | 1     | CUT2 HUMAN  |
| 29         | 66.5  | 15.0        | 3149   | 1     | TGCU1 EBV   |
| 30         | 66    | 14.9        | 551    | 1     | PODX RABIT  |
| 31         | 66    | 14.9        | 621    | 1     | ILVB MYCAV  |
| 32         | 65.5  | 14.8        | 325    | 1     | APA MYCBO   |
| 33         | 65.5  | 14.8        | 325    | 1     | APA_MYCTU   |

|    |      |      |      |   |            |
|----|------|------|------|---|------------|
| 34 | 65.5 | 14.8 | 329  | 1 | EXBB PSEPU |
| 35 | 65.5 | 14.8 | 812  | 1 | ERN3 EBV   |
| 36 | 65.5 | 14.8 | 1004 | 1 | SAL2 MOUSE |
| 37 | 65.5 | 14.8 | 1093 | 1 | AFI7 HUMAN |
| 38 | 65.5 | 14.8 | 3511 | 1 | MY15 MOUSE |
| 39 | 65   | 14.7 | 209  | 1 | CXX1 HUMAN |
| 40 | 65   | 14.7 | 328  | 1 | ZIPA ECOLI |
| 41 | 65   | 14.7 | 328  | 1 | ZIPA ECOLI |
| 42 | 65   | 14.7 | 361  | 1 | IF35 MOUSE |
| 43 | 65   | 14.7 | 427  | 1 | TER4 HUMAN |
| 44 | 65   | 14.7 | 724  | 1 | P85B BOVIN |
| 45 | 64.5 | 14.6 | 346  | 1 | SELV HUMAN |

## ALIGNMENTS

|          |   |                                   |      |         |
|----------|---|-----------------------------------|------|---------|
| RESULT 1 | IEXL1_HUMAN   | STANDARD;                         | PRT; | 156 AA. |
| ID       | AC  | P46695; Q92691; Q93044;           |      |         |
| DT       | 01-NOV-1995   | (Rel. 32, Created)                |      |         |
| DT       | 15-JUL-1998   | (Rel. 36, Last sequence update)   |      |         |
| DT       | 15-MAR-2004   | (Rel. 43, Last annotation update) |      |         |
| DE       | Radiation-inducible immediate-early gene IEX-1 (Immediate early protein GLY96) (Immediate early response 3 protein) (PACAP-responsive gene 1 protein) (PRGI protein) (Differentiation-dependent gene 2 protein) (DIF-2 protein).  |                                   |      |         |
| DE       | IEX3 OR IEX1 OR PRGI OR DIF2.   |                                   |      |         |
| OS       | Homo sapiens (Human).   |                                   |      |         |
| OC       | Rukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |                                   |      |         |
| OC       | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |                                   |      |         |
| OX       | NCBI_TaxID=9606;  |                                   |      |         |
| RN       | [1]_TaxID=9606;   |                                   |      |         |
| RP       | SEQUENCE FROM N.A.  |                                   |      |         |
| RC       | TISSUE=Placenta;  |                                   |      |         |
| RX       | MEDLINE=96181295; PubMed=8603392;   |                                   |      |         |
| RA       | Kondratyev A.D., Chung K.-N., Jung M.O.;  |                                   |      |         |
| RT       | Identification and characterization of a radiation-inducible glycosylated human early-response gene.;   |                                   |      |         |
| RT       | Cancer Res. 56:1498-1502(1996).   |                                   |      |         |
| RL       | [2]   |                                   |      |         |
| RN       | SEQUENCE FROM N.A.  |                                   |      |         |
| RP       | MEDLINE=96221139; PubMed=8653710;   |                                   |      |         |
| RX       | Schaefer H., Trauzold A., Siegel E.G., Folsch U.R., Schmidt W.E.;   |                                   |      |         |
| RA       | PRGI: a novel early-response gene transcriptionally induced by pituitary adenylate cyclase activating polypeptide in a pancreatic carcinoma cell line.;   |                                   |      |         |
| RT       | Cancer Res. 56:2641-2641(1996).   |                                   |      |         |
| RL       | [3]   |                                   |      |         |
| RN       | SEQUENCE FROM N.A.  |                                   |      |         |
| RP       | MEDLINE=97339426; PubMed=9196025;   |                                   |      |         |
| RX       | Pietzsch A., Buechler C., Aslanidis C., Schmitz G.;   |                                   |      |         |
| RA       | Identification and characterization of a novel monocyte/macrophage differentiation-dependent gene that is responsive to lipopolysaccharide, ceramide, and lysophosphatidylcholine.;   |                                   |      |         |
| RT       | Biochem. Biophys. Res. Commun. 235:4-9(1997).   |                                   |      |         |
| RL       | [4]   |                                   |      |         |
| RN       | SEQUENCE FROM N.A.  |                                   |      |         |
| RP       | Shina S., Tamiya G., Oka A., Inoko H.;  |                                   |      |         |
| RX       | Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.;   |                                   |      |         |
| RA       | Submitted (SBP-1999) to the EMBL/GenBank/DBJ databases.   |                                   |      |         |
| RL       | [5]   |                                   |      |         |
| RN       | SEQUENCE FROM N.A.  |                                   |      |         |
| RP       | TISSUE=Cervix, and Skin;  |                                   |      |         |
| RC       | MEDLINE=22386257; PubMed=12477932;  |                                   |      |         |
| RX       | Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Frange C., |                                   |      |         |

|        |              |
|--------|--------------|
| Q05605 | pseudomonas  |
| P12977 | epstein-bar  |
| Q9QX96 | mus musculus |
| P55138 | homo sapien  |
| Q9GZ24 | mus musculus |
| O15255 | homo sapien  |
| Q8X492 | escherichia  |
| P77173 | escherichia  |
| Q9dch4 | mus musculus |
| Q15561 | homo sapien  |
| P23726 | bos taurus   |
| P59797 | homo sapien  |

```
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gurnaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Sherchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- INDUCTION: By radiation, TPA, okadaic acid and TNF-alpha.
CC -!- PTM: Glycosylated.
CC -!- SIMILARITY: Belongs to the IER3 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; S81914; AAB36278.1; -
DR EMBL; X96438; CAA65304.1; -
DR EMBL; Y14551; CAA74886.1; -
DR EMBL; AP00512; BAB63319.1; -
DR EMBL; BC000844; AAH00844.1; -
DR EMBL; BC005080; AAH05080.1; -
DR Gene; HGNC:15392; IER3.
DR MIN; 602996; -
DR GO; GO:0008189; P:apoptosis inhibitor activity; TAS.
DR GO; GO:0006916; P:anti-apoptosis; TAS.
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR GO; GO:0007345; P:embryogenesis and morphogenesis; TAS.
KW Glycoprotein; Transmembrane; Signal-anchor.
FT DOMAIN 1 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 99 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT
FT DOMAIN 100 156 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLNAC. .) (POTENTIAL).
FT CONFLICT 54 54 A -> G (IN REF. 1).
FT CONFLICT 106 106 P -> R (IN REF. 1).
SQ SEQUENCE 156 AA; 16928 MW; 83C06116C80B9240 CRC64;

Query Match 100.0%; Score 442; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.7e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRQLVPEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPDAPNAASLAPTPVSPVLE 60
Db 71 VRQLVPEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPDAPNAASLAPTPVSPVLE 130
QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86
Db 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 2
ID TEXT_PANTR STANDARD; PRT; 156 AA.
AC Q7YR42;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Radiation-inducible immediate-early gene IEX-1 (Immediate early
DE protein Gly96) (Immediate early response 3 protein).
GN IER3 OR IEX1.
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Pan troglodytes (Chimpanzee).
CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22709134; PubMed=12799463;
RA Anzai T., Shiina T., Kimura N., Yanagiya K., Kohara S., Shigenari A.,
RA Yamagata T., Kulski J.K., Naruse T.K., Fujimori Y., Fukuzumi Y.,
RA Yamazaki M., Tashiro H., Iwamoto C., Uehara Y., Imanishi T.,
RA Meyer A., Ikeo K., Gojobori T., Bahram S., Inoko H.;
RT "Comparative sequencing of human and chimpanzee MHC class I regions
RT unveils insertions/deletions as the major path to genomic
RT divergence."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7708-7713(2003).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
CC -!- PTM: Glycosylated (By similarity).
CC -!- SIMILARITY: Belongs to the IER3 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AB100084; BAC78173.1; -
KW Glycoprotein; Transmembrane; Signal-anchor.
FT DOMAIN 1 82 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 83 99 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT
FT DOMAIN 100 156 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 133 133 N-LINKED (GLNAC. .) (POTENTIAL).
SQ SEQUENCE 156 AA; 16929 MW; 83C067CDCAC09650 CRC64;

Query Match 100.0%; Score 442; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 1.7e-36;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRQLVPEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPDAPNAASLAPTPVSPVLE 60
Db 71 VRQLVPEPNPAKRLFLLLITVFCQILMAEGVPAPLPDPDAPNAASLAPTPVSPVLE 130
QY 61 PFNLTSEPSDYALDLSTFLQHPAAF 86
Db 131 PFNLTSEPSDYALDLSTFLQHPAAF 156

RESULT 3
ID TEXT_MOUSE STANDARD; PRT; 153 AA.
AC P46694;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Radiation-inducible immediate-early gene IEX-1 (Immediate early
DE protein Gly96) (Immediate early response 3 protein).
GN IER3 OR IEX1 OR Gly96.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Charles C.H., Yoon J.K., Sims J.S., Lau L.F.;
RT "Genomic structure, cDNA sequence, and expression of gly96, a growth
RT factor-inducible immediate-early gene encoding a short-lived
RT glycosylated protein."
RL Oncogene 8:797-801(1993).
CC -!- FUNCTION: Not known; expressed during the G0-G1 transition of the
CC cell cycle.
CC -!- SUBCELLULAR LOCATION: Type II membrane protein (Potential).
```

```

CC -!- TISSUE SPECIFICITY: Expressed predominantly in the lung, testes
CC and the uterus.
CC -!- INDUCTION: By serum growth factors and TPA.
CC -!- PM: Glycosylated.
CC -!- SIMILARITY: Belongs to the IER3 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X67644; -; NOT_ANNOTATED_CDS.
CC FIC; S33363; S33363.
CC MGD; MGI:10484; Ier3.
KW Glycoprotein; Transmembrane; Signal-anchor.
FT DOMAIN 1 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 102 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT DOMAIN 103 153 EXTRACELLULAR (POTENTIAL).
FT CARBOHYD 137 137 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 153 AA; 16875 MW; F4429E3B0120DE11 CRC64;

Query Match 61.8%; Score 273; DB 1; Length 153;
Best Local Similarity 68.2%; Pred. No. 5.5e-20;
Matches 58; Conservative 5; Mismatches 14; Indels 8; Gaps 2;

QY 1 VRQQLPVEPNKRLFLLLITVFCQILMAEGVPALPPEDAPNAASLAPTYS----56
DB 71 VRQQLPVEPNKRLFLLLITVFCQILMAEGVPALPPEDAPNAASLAPTYS----56
QY 57 --PVLSPFNLSPESDVALDSTEL 79
DB 129 APPVLEPLNLTSSESDYALDKRAFL 153

RESULT 4
SOS_DROME STANDARD; PRT; 1595 AA.
AC P26675;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Son of sevenless protein.
GN SOS.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oregon-R;
RX MEDLINE=92141920; PubMed=1736363;
RA Bonfini L., Karlovich C.A., Dasgupta C., Banerjee U.;
RT "The Son of sevenless gene product: a putative activator of Ras.";
RL Science 255:603-606(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92034991; PubMed=1934068;
RA Simon M.A., Bowtell D.D.L., Dodson G.S., Lavery T.R., Rubin G.M.;
RT "Ras1 and a putative guanine nucleotide exchange factor perform
RT crucial steps in signaling by the sevenless protein tyrosine
RT kinase.".;
RL Cell 67:701-716(1991).
CC -!- FUNCTION: Promotes the exchange of Ras-bound GTP by GTP. SOS is
CC implicated in neuronal development.
CC -!- SUBUNIT: May form a complex with sevenless and DRK.
CC -!- SIMILARITY: Contains 1 DBL-homology (DRH) domain.
CC -!- SIMILARITY: Contains 1 N-terminal Ras-GEF domain.
CC -!- SIMILARITY: Contains 1 PH domain.

```

```

CC -!- SIMILARITY: Contains 1 Ras-GEF domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M83931; AAB04680.1; -.
CC EMBL; M77501; AAA28904.1; -.
CC PDB; 1AZE; 18-MAY-99.
CC FlyBase; FBgn0001965; Sos.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR007124; Hist_TAF.
CC InterPro; IPR001849; PH.
CC InterPro; IPR008937; Ras_GEF.
CC InterPro; IPR000651; RasGEFN.
CC InterPro; IPR001895; RasGEF_CDC25.
CC InterPro; IPR000219; RhoGEF.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEFN; 1.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00147; RasGEF; 1.
CC SMART; SM00229; RasGEFN; 1.
CC SMART; SM00325; RhoGEF; 1.
CC PROSITE; PS00741; DH_1; FALSE_NEG.
CC PROSITE; PS0010; DH_2; 1.
CC PROSITE; PS00720; RasGEF; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC PROSITE; PS50009; RasGEF_CAT; 1.
CC PROSITE; PS50212; RasGEF_NTER; 1.
KW Guanine-nucleotide releasing factor; Neurogenesis; 3D-structure.
FT DOMAIN 636 791 N-TERMINAL RAS-GEF.
FT DOMAIN 828 1065 RAS-GEF.
FT DOMAIN 15 51 GLY-RICH.
FT DOMAIN 247 433 DH.
FT DOMAIN 479 587 PH.
FT DOMAIN 1511 1516 GLN-RICH.
FT DOMAIN 1525 1541 HIS-RICH.
FT CONFLICT 232 243 TSCVPCHPPRS -> HILSPSLPAQR
(IN REF. 2)
FT CONFLICT 1462 1462 V -> P (IN REF. 2).
SQ SEQUENCE 1595 AA; 177837 MW; 33AE31F0767A219F CRC64;

Query Match 17.8%; Score 78.5; DB 1; Length 1595;
Best Local Similarity 28.3%; Pred. No. 7.2;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRQQLPVE-----EPNPKRLFLLLITVFCQILMAEGVPALPPEDAPNAASLAPT 53
DB 1422 IRRNSAIEKRAAATSQPNQAAGPISITLVTSQAVATDEVLPISP-----AASSTT 1476
QY 54 --PVSPLVPP--NLTSEP-----SDYALDSTELQO---HPAAP 85
DB 1477 TSPLTAMSPMGPNIPSPVSTSSSYAHQLMRQOQQOQTHPIY 1522

RESULT 5
MLL2_HUMAN STANDARD; PRT; 5262 AA.
AC Q14686; Q14687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related
DE protein).
GN MLL2 CR ALL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```





```

FT PEEPLSPQAEGRHLSQPELHLSPCTREPHLSVPVREPC
FT LSPQEPESHLSPOSSEPCLSRPRSHLSPELEKPLSPRP
FT EKPEEFGQCAPAEELPLFPFPPGSPSLGLGEPALSPGE
FT PPLSPLEPLSPSPGSPSLSPQLMPDPLPPLSPITAA
FT A (in isoform 2).
FT /FTid=VSP_008559.
FT VARSPLIC 1454 1454 E -> EGET (in isoform 3).
FT /FTid=VSP_008560.
FT VARIANT 4949 4949 R -> H (in GSNP:3782356).
FT /FTid=VAR_017115.
SQ SEQUENCE 5262 AA; 564171 MW; 26B7C74CAD417B44 CRC64;

Query Match 17.6%; Score 78; DB 1; Length 5262;
Best Local Similarity 30.4%; Pred. No. 30;
Matches 21; Conservative 12; Mismatches 28; Indels 8; Gaps 2;

QY 1 VRRQLVPEEPNPAKRLFLILLITVFCQILMAEKGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 4604 LKQSPAPERTQHYTYNVNSLDVRQL-----SAPPEEPSPSPSPASPTPE 4656

QY 61 PF-NLTSPP 68
Db 4657 PLVELPTEP 4665

RESULT 6
GSQC ERWCH STANDARD; PRT; 272 AA.
AC P31698;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE General secretion pathway protein C (pectic enzymes secretion protein
DE outC).
GN OUTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937.
RX MEDLINE=93054355; PubMed=1429461;
RA Lindeberg M., Collmer A.;
RT "Analysis of eight out genes in a cluster required for pectic enzyme
RT secretion by Erwinia chrysanthemi: sequence comparison with secretion
RT genes from other Gram-negative bacteria.";
RL J. Bacteriol. 174:7385-7397(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: Inner membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65265; CAA46369.1; -.
CC PIR; S28013; S28013.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001639; T2SP_C.
CC Pfam; PF00595; PDZ; 1.
CC PRINTS; PR00810; BCTERIALGSPC.
CC PROSITE; PS01141; T2SP_C; 1.
CC Transport; transmembrane; Inner membrane.
CC DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 17 35 POTENTIAL.
CC DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
CC EMBL; L02214; AAA24830.1; -.
CC PIR; A47021; A47021.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001639; T2SP_C.
CC Pfam; PF00595; PDZ; 1.
CC PRINTS; PR00810; BCTERIALGSPC.
CC PROSITE; PS01141; T2SP_C; 1.
CC Transport; transmembrane; Inner membrane.
CC DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 17 35 POTENTIAL.
CC DOMAIN 36 272 PERIPLASMIC (POTENTIAL).

```

```

SQ SEQUENCE 272 AA; 30301 MW; 064C7311P9714405 CRC64;

Query Match 16.6%; Score 73.5; DB 1; Length 272;
Best Local Similarity 30.1%; Pred. No. 3.2;
Matches 22; Conservative 13; Mismatches 25; Indels 13; Gaps 3;

QY 4 QLPVEEPNPAKRLFLILLITVFCQ---ILMAEKGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 5 KLPPSPSPVIRRLFYLLMLLFCQQLAMIFWRVGLP-----DNSPVASVOITPAQARQQ 58
QY 61 PFNLTSPPSDVAL 73
Db 59 PVTL----NDFTL 67

RESULT 7
GSQC ERWCH STANDARD; PRT; 272 AA.
AC Q01564;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE General secretion pathway protein C (pectic enzymes secretion protein
DE outC).
GN OUTC.
OS Erwinia chrysanthemi.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3937.
RX MEDLINE=93056427; PubMed=1453958;
RA Condemine G., Dorel C., Hugouvieux-Cotte-Pattat N., Robert-Baudouy J.;
RT "Some of the out genes involved in the secretion of pectate lyases in
RT Erwinia chrysanthemi are regulated by kdgr.";
RL Mol. Microbiol. 6:3199-3211(1992).
CC -!- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS. REQUIRED FOR THE TRANSLLOCATION OF THE
CC MULTIPLE PECTIC ENZYMES.
CC -!- SUBCELLULAR LOCATION: Inner membrane (Probable).
CC -!- SIMILARITY: BELONGS TO THE EXEC/PULC/OUTC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X65265; CAA46369.1; -.
CC PIR; S28013; S28013.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001639; T2SP_C.
CC Pfam; PF00595; PDZ; 1.
CC PRINTS; PR00810; BCTERIALGSPC.
CC PROSITE; PS01141; T2SP_C; 1.
CC Transport; transmembrane; Inner membrane.
CC DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 17 35 POTENTIAL.
CC DOMAIN 36 272 PERIPLASMIC (POTENTIAL).
CC SEQUENCE 272 AA; 30162 MW; 0E2C1A952B42605A CRC64;

Query Match 16.4%; Score 72.5; DB 1; Length 272;
Best Local Similarity 30.1%; Pred. No. 4;
Matches 22; Conservative 16; Mismatches 22; Indels 13; Gaps 4;

QY 4 QLPVEEPNPAKRLFLILLITVFCQ---ILMAEKGVPAPLPEDAPNAASLAPTPVSPVLE 60
Db 5 KLPPSPSPVIRRLFYLLMLLFCQQLAMIFWRVGLP-----PDNAP-VSSVOITPAQARQQ 58
QY 61 PFNLTSPPSDVAL 73

```





RX MEDLINE=93214031; PubMed=8384998;  
 RA Russell S.D., Roth G.J.;  
 RA "pseudo-von Willebrand disease: a mutation in the platelet  
 RT glycoprotein Ib alpha gene associated with a hyperactive surface  
 RT receptor";  
 RL Blood 81:1787-1791(1993).  
 RL [17]  
 RP VARIANT BSS LEU-195 DEL.  
 RP MEDLINE=95179321; PubMed=7873390;  
 RA de la Salle C., Baas M.-J., Lanza F., Schwartz A., Hanau D.,  
 RA Chevalier J., Gachet C., Briquel M.-E., Cazenave J.-P.;  
 RA "A three-base deletion removing a leucine residue in a leucine-rich  
 RT repeat of platelet glycoprotein Ib alpha associated with a variant of  
 RT Bernard-Soulier syndrome (Nancy I).";  
 RL Br. J. Haematol. 89:386-396(1995).  
 RL [18]  
 RP VARIANT BSS ARG-81.  
 RP MEDLINE=98303759; PubMed=9639514;  
 RA Kenny D., Jonsson O.G., Morateck P.A., Montgomery R.R.;  
 RA "Naturally occurring mutations in glycoprotein Ib alpha that result in  
 RT defective ligand binding and synthesis of a truncated protein.";  
 RL Blood 92:175-183(1998).  
 RL [19]  
 RP VARIANTS HIS-72 AND MET-161.  
 RP MEDLINE=99318093; PubMed=1091209;  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemes J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.-Q.,  
 RA Lander E.S.;  
 RA "Characterization of single-nucleotide polymorphisms in coding regions  
 RT of human genes.";  
 RL Nat. Genet. 22:231-238(1999).  
 RL [20]  
 RP ERATUM.  
 RA Cargill M., Altshuler D., Ireland J., Sklar P., Ardlie K., Patil N.,  
 RA Shaw N., Lane C.R., Lim E.P., Kalyanaram N., Nemes J., Ziaugra L.,  
 RA Friedland L., Rolfe A., Warrington J., Lipshutz R., Daley G.-Q.,  
 RA Lander E.S.;  
 RL Nat. Genet. 23:373-373(1999).  
 RL [21]  
 RP VARIANT BSS PRO-145.  
 RP MEDLINE=99189763; PubMed=10089893;  
 RA Koskela S., Partanen J., Salmi T.T., Kekomaki R.;  
 RA "Molecular characterization of two mutations in platelet glycoprotein  
 RT (GP) Ib alpha in two Finnish Bernard-Soulier syndrome families.";  
 RL Eur. J. Haematol. 62:160-168(1999).  
 RL CC -!- FUNCTION: GP-Ib, a surface membrane protein of platelets,  
 CC participates in the formation of platelet plugs by binding to the  
 CC A1 domain of von Willebrand factor, which is already bound to the  
 CC subendothelium.  
 CC CC -!- SUBUNIT: Heterodimer composed of GP-Ib alpha and beta; disulfide  
 CC linked. GP-IX is complexed with the GP-Ib heterodimer via a non  
 CC covalent linkage. Interacts with FLNB.  
 CC CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC CC -!- PTM: Glycosylation, which is approximately coextensive with the  
 CC extracellular part of the molecule, is cleaved off by calpain  
 CC during platelet lysis.  
 CC CC -!- POLYMERISM: Position 161 is associated with platelet-specific  
 CC alloantigen Siba. Siba(-) has Thr-161 and Siba(+) has Met-161.  
 CC CC Siba is involved in neonatal alloimmune thrombocytopenia (NATP).  
 CC CC -!- DISEASE: Defects in GP1BA are one of the causes of Bernard-Soulier  
 CC syndrome (BSS) [MIM:153670]. BSS patients have unusually large  
 CC platelets and have a clinical bleeding tendency.  
 CC CC -!- DISEASE: Defects in GP1BA are one of the causes of von Willebrand  
 CC disease (vWD) [MIM:177820]; also known as platelet-type von  
 CC Willebrand disease or pseudo-von Willebrand disease [pseudo-vWD].  
 CC CC This autosomal dominant bleeding disorder is caused by an  
 CC increased affinity of GP-Ib for soluble vWF resulting in impaired  
 CC hemostatic function due to the removal of vWF from the  
 CC circulation.



```

FT DOMAIN 2251 2259 POLY-PRO.
FT VARSPLIC 532 582 VARSRVIKTTPRRFWEDEPPKPKVSVLRPIITSPVP
FT VPQEPAPVPS -> PLSQSLIPMTQLSLSGWAAPTTS
FT ACIDSLPWSPLLRPRCPUTGLQL (in isoform 2).
FT /FTID=VSP_006668.
FT Missing (in isoform 2).
FT /FTID=VSP_006669.
FT K -> E (IN REF. 6).
FT CONFLICT 834 834
FT CONFLICT 941 941
FT CONFLICT 1317 1317 S -> Y (IN REF. 6).
FT CONFLICT 1362 1362 H -> Q (IN REF. 6).
FT CONFLICT 1438 1438 D -> N (IN REF. 6).
FT CONFLICT 1918 1920 PLA -> QTR (IN REF. 5; RAH093337).
FT CONFLICT 2541 2543 DEE -> ARG (IN REF. 5; AAH07353).
FT CONFLICT 2622 2622 D -> H (IN REF. 6).
SQ SEQUENCE 2715 AA; C0615B981BBEB7BF CRC64;

Query Match 16.0%; Score 70.5; DB 1; Length 2715;
Best Local Similarity 34.5%; Pred. No. 78;
Matches 19; Conservative 8; Mismatches 27; Indels 1; Gaps 1;

QY 8 EEPNPAKLLFLITVFCQILMAEAGVPAPLPEDAPNAAFLPT-FVSPVLEP 61
Db 2182 EPPKPAISKILVNKLQGVKMGAGEPVPVVKQPLPPTISPTAFTSWTLPP 2236

RESULT 12
GNDS HUMAN
ID AC GNDS HUMAN STANDARD; PRT; 914 AA.
IC Q12967; Q9HAX7; Q9HAX1; Q9HCW1;
DI 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ral guanine nucleotide dissociation stimulator (RalGDS). (RalGDS).
GN RALGDS OR RGF.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98032487; PubMed=9365783;
RA Humphrey D., Kwiatkowska J., Henske E.P., Haines J.L., Halley D.,
RA van Slegtenhorst M., Kwiatkowski D.J.;
RT "Cloning and evaluation of RALGDS as a candidate for the tubercous
RT sclerosis gene TSC1."
RL Ann. Hum. Genet. 61:299-305(1997).
RN [2]
RP SEQUENCE OF 587-914 FROM N.A.
RX MEDLINE=95062211; PubMed=7972015;
RA Hofer F., Fields S., Schneider C., Martin G.S.;
RT "Activated Ras interacts with the Ral guanine nucleotide dissociation
RT stimulator."
RL Proc. Natl. Acad. Sci. U.S.A. 91:11089-11093(1994).
RN [3]
RP STRUCTURE BY NMR OF 788-884.
RX MEDLINE=97448669; PubMed=9020994;
RA Geyer M., Herrmann C., Wohlgemuth S., Wittinghofer A., Kalbitzer H.R.;
RT "Structure of the Ras-binding domain of RalGAP and implications for
RT Ras binding and signaling."
RL Nat. Struct. Biol. 4:694-699(1997).
RN [4]
RP STRUCTURE BY NMR OF 771-886.
RA Mueller T.D., Handel L., Schmieder P., Oschkinat H.;
RT "High-resolution structure of the RA-domain of human RALGDS and a
RT dynamics study of its binding loop to ras."
RL Submitted (MAR-1999) to the PDB data bank.
CC -!- FUNCTION: Stimulates the dissociation of GTP from the Ras-related
CC Rala and Ralb GTPases which allows GTP binding and activation of
CC the GTPases. Interacts and acts as an effector molecule for R-Ras,
CC H-Ras, K-Ras, and Rap.
CC -!- DOMAIN: The Ras-associating domain interacts with Ras.

```

```

CC -!- SIMILARITY: Contains 1 N-terminal Ras-GEF domain.
CC -!- SIMILARITY: Contains 1 Ras-associating domain.
CC -!- SIMILARITY: Contains 1 Ras-GEF domain.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF295773; AAG02122.1; -.
CC EMBL; AF295775; AAG10221.1; -.
CC EMBL; AF295778; AAG10221.1; JOINED.
CC EMBL; AF295780; AAG10225.1; -.
CC EMBL; U14417; AAS2360.1; -.
CC PIR; I38953; I38853.
CC PDB; 2RGF; 04-MAR-98.
CC PDB; 1PAX; 19-MAR-98.
CC Genew; HGNC:9842; RALGDS.
CC MIM; 601619; -.
CC GO; GO:0005083; F:small GTPase regulatory/interacting protein. . .; IDA.
CC GO; GO:0007264; P:small GTPase mediated signal transduction; NAS.
CC InterPro; IPR000159; RA domain.
CC InterPro; IPR008937; Ras GEF.
CC InterPro; IPR000651; RasGEFN.
CC InterPro; IPR001895; RasGRP_CDC25.
CC Pfam; PF00789; RA; 1.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEFN; 1.
CC SMART; SM00314; RA; 1.
CC SMART; SM00147; RasGEF; 1.
CC SMART; SM00229; RasGEFN; 1.
CC PROSITE; PS00720; RasGEF; 1.
CC PROSITE; PS50200; RA; 1.
CC PROSITE; PS50009; RasGEF CAT; 1.
CC PROSITE; PS50212; RasGEF NTER; 1.
CC Guanine-nucleotide releasing factor; 3D-structure.
CC DOMAIN 112 249 N-TERMINAL RAS-GEF.
CC DOMAIN 386 648 RAS-GEF.
CC DOMAIN 798 885 RAS-ASSOCIATING.
CC DOMAIN 714 717 POLY-SER.
CC DOMAIN 753 765 POLY-SER.
CC DOMAIN 839 844 POLY-SER.
CC STRAND 792 806 POLY-GLU.
CC STRAND 815 820
CC TURN 821 822
CC HELIX 825 835
CC TURN 836 839
CC TURN 843 844
CC STRAND 846 851
CC TURN 854 855
CC STRAND 857 860
CC TURN 864 865
CC HELIX 867 871
CC TURN 872 873
CC STRAND 877 882
CC SSSEQUENCE 914 AA; 100606 MW; EA5A5CF25AF3D523 CRC64;

Query Match 15.7%; Score 69.5; DB 1; Length 914;
Best Local Similarity 37.4%; Pred. No. 30;
Matches 34; Conservative 7; Mismatches 29; Indels 21; Gaps 8;

QY 5 LPVEENPAKLLFLITVFCQILMAEAGVPAPLP-PEDAPNAA-----SLAPTEVSPV 58
Db 264 VVALKPTPE---LEALPTA-----RAPSPVAPAPPEPAPFPAPGSELEVAPAP-AP 314
QY 59 L-----EP-FNLTSEPSDYALDLTLFQQHPA 84
Db 315 LQAPEPAVGLSESAPAP-ALELEPAPEQDPA 344

```







RT "Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [5]  
RP SEQUENCE OF 1-358 FROM N.A., AND SUBCELLULAR LOCATION.  
RC TISSUE=Fetal brain;  
RX MEDLINE=99068504; PubMed=9853615;  
RA Ueki N., Oda T., Kondo M., Yano K., Noguchi T., Muramatsu M.-A.;  
RT "Selection system for genes encoding nuclear-targeted proteins.";  
RL Nat. Biotechnol. 16:1338-1342(1998).  
CC -!- FUNCTION: Could play a role in constitutive transport element  
CC (CTE)-mediated gene expression. Does not seem to be implicated in  
CC the binding of regulatory subunit II of PKA.  
CC -!- SUBUNIT: Binds to the C-terminal of RNA helicase A.  
CC -!- SUBCELLULAR LOCATION: Nuclear at steady state but shuttles between  
CC the nucleus and cytoplasm.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AB025905; BAA85003.1; -.  
DR EMBL; AJ243467; CAB65092.1; -.  
DR EMBL; AF799414; AAF86048.1; -.  
DR EMBL; EC000713; AAH00713.1; -.  
DR EMBL; AB015332; BAA34791.1; ALT\_INIT.  
DR GO; GO:0005634; C:nucleus; TAS.  
DR GO; GO:0017151; F:DEAD/H-box RNA helicase binding; TAS.  
DR InterPro; IPR007071; AkAP95.  
DR InterPro; IPR007087; Znf\_C2H2.  
DR Pfam; PF04988; AKAP95; 1.  
DR SMART; SM00355; Znf\_C2H2; 1.  
KW Nuclear protein; Zinc-finger.  
FT DOMAIN 274 279 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT DOMAIN 362 364 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
FT ZN\_FING 391 413 C2H2-TYPE.  
FT ZN\_FING 484 507 C2H2-TYPE.  
FT DOMAIN 41 52 TYR/GLY-RICH.  
FT DOMAIN 602 608 PRO-RICH.  
FT DOMAIN 589 597 PRO-RICH.  
FT CONFLICT 100 100 D -> N (IN REF. 3).  
FT CONFLICT 189 189 S -> N (IN REF. 3).  
FT CONFLICT 351 358 ALTYQDEN -> EFSGAWC (IN REF. 4).  
FT CONFLICT 458 458 Q -> H (IN REF. 1).  
FT CONFLICT 554 596 EEKQEEAGGGALDEGAAGISGAGGVPAQPPVPE  
FT PA -> RRRSRRLRAVFWTRGRKQGFRAQACRRS  
FT LPPQSQP (IN REF. 3).  
FT CONFLICT 610 640 EEEGAVPLGGALGRLDVEDDER -> GGGGGR  
FT RGCWEGRCNARSASASTWTTTKK (IN REF. 3).  
SQ SEQUENCE 646 AA; 71640 MW; CBC265BF25996BA2 CRC64;  
Query Match 15.6%; Score 69; DB 1; Length 646;  
Best Local Similarity 62.5%; Pred. No. 23;  
Matches 15; Conservative 1; Mismatches 6; Indels 2; Gaps 1;  
QY 33 EGYPA--PLPPEDAPNAASLAPTP 54  
| | | | | : | | | | | | | | | |  
Db 583 EGVFAQFPVPEPAPGAVSPPPP 606

Search completed: February 25, 2004, 05:34:03  
Job time : 15 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 25, 2004, 05:32:32 ; Search time 39 Seconds  
(without alignments)  
695.758 Million cell updates/sec

Title: US-08-799-910-10\_COPY\_71\_156

Perfect score: 442

Sequence: 1 VRRQLPVEENPAKLLFLL.....EPSDYALDLSFLQQHPAAF 86

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 segs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:\*
- 1: sp archaea:\*
  - 2: sp bacteria:\*
  - 3: sp fungi:\*
  - 4: sp human:\*
  - 5: sp invertebrate:\*
  - 6: sp mammal:\*
  - 7: sp mhc:\*
  - 8: sp organelle:\*
  - 9: sp phage:\*
  - 10: sp plant:\*
  - 11: sp rodent:\*
  - 12: sp virus:\*
  - 13: sp vertebrate:\*
  - 14: sp unclassified:\*
  - 15: sp virus:\*
  - 16: sp bacteriaph:\*
  - 17: sp archaea:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length | ID        | Description         |
|------------|-------|-------------|--------|-----------|---------------------|
| 1          | 442   | 100.0       | 156    | 6 Q7YR42  | Q7YR42 pan troglod  |
| 2          | 442   | 100.0       | 193    | 4 Q75353  | Q75353 homo sapien  |
| 3          | 312   | 70.6        | 160    | 11 Q91V25 | Q91V25 mus musculu  |
| 4          | 84.5  | 19.1        | 335    | 4 Q96CZ8  | Q96CZ8 homo sapien  |
| 5          | 84.5  | 19.1        | 926    | 4 Q9H0K1  | Q9H0K1 homo sapien  |
| 6          | 81.5  | 18.4        | 102    | 5 Q9T2D3  | Q9T2D3 caenorhabdi  |
| 7          | 81.5  | 18.4        | 452    | 5 Q9JAS5  | Q9JAS5 euploties cr |
| 8          | 81.5  | 18.4        | 468    | 5 Q9XY65  | Q9XY65 euploties cr |
| 9          | 81.5  | 18.4        | 931    | 11 Q8CFH6 | Q8CFH6 mus musculu  |
| 10         | 80.5  | 18.2        | 302    | 12 Q9QM69 | Q9QM69 avian adeno  |
| 11         | 80.5  | 18.2        | 343    | 5 Q8ITK2  | Q8ITK2 drosophila   |
| 12         | 79.5  | 18.0        | 343    | 5 Q8ITK1  | Q8ITK1 drosophila   |
| 13         | 79.5  | 18.0        | 343    | 5 Q810P7  | Q810P7 drosophila   |
| 14         | 79.5  | 18.0        | 343    | 5 Q810E7  | Q810E7 drosophila   |
| 15         | 79.5  | 18.0        | 372    | 5 Q95T18  | Q95T18 drosophila   |
| 16         | 79.5  | 18.0        | 1596   | 5 Q9VJTV9 | Q9VJTV9 drosophila  |

|    |      |      |      |           |                    |
|----|------|------|------|-----------|--------------------|
| 17 | 77.5 | 17.5 | 318  | 16 Q88BE8 | Q88BE8 pseudomonas |
| 18 | 77   | 17.4 | 1568 | 4 Q95785  | Q95785 homo sapien |
| 19 | 77   | 17.4 | 2602 | 5 Q9VZ45  | Q9VZ45 drosophila  |
| 20 | 76.5 | 17.3 | 229  | 16 Q8ENW9 | Q8ENW9 xanthomonas |
| 21 | 76   | 17.2 | 148  | 16 Q9JZR7 | Q9JZR7 neisseria m |
| 22 | 76   | 17.2 | 458  | 11 Q80V59 | Q80V59 mus musculu |
| 23 | 76   | 17.2 | 801  | 5 Q9W3K6  | Q9W3K6 drosophila  |
| 24 | 76   | 17.2 | 849  | 11 Q80V17 | Q80V17 mus musculu |
| 25 | 76   | 17.2 | 1677 | 5 Q00S05  | Q00S05 leishmania  |
| 26 | 75.5 | 17.1 | 1136 | 3 Q9HGK8  | Q9HGK8 tuber borch |
| 27 | 75   | 17.0 | 982  | 11 Q7TSJ4 | Q7TSJ4 mus musculu |
| 28 | 74   | 16.7 | 114  | 12 Q9YIQ8 | Q9YIQ8 avian adeno |
| 29 | 74   | 16.7 | 955  | 11 Q88287 | Q88287 mus musculu |
| 30 | 74   | 16.7 | 1180 | 5 Q7YU19  | Q7YU19 drosophila  |
| 31 | 74   | 16.7 | 1306 | 5 Q9VH05  | Q9VH05 drosophila  |
| 32 | 74   | 16.7 | 1561 | 11 Q88286 | Q88286 mus musculu |
| 33 | 73.5 | 16.6 | 276  | 16 Q9JSG9 | Q9JSG9 bradyrhizob |
| 34 | 73.5 | 16.6 | 406  | 2 Q8GAM6  | Q8GAM6 arthrobacte |
| 35 | 73.5 | 16.6 | 925  | 12 Q69138 | Q69138 human herpe |
| 36 | 73   | 16.5 | 708  | 16 Q7WAY9 | Q7WAY9 bordetella  |
| 37 | 73   | 16.5 | 743  | 16 Q7WFE3 | Q7WFE3 bordetella  |
| 38 | 73   | 16.5 | 1029 | 12 Q9YMX0 | Q9YMX0 lymantria d |
| 39 | 72.5 | 16.4 | 474  | 5 Q86154  | Q86154 dictyosteli |
| 40 | 72   | 16.3 | 659  | 5 Q9VQ22  | Q9VQ22 drosophila  |
| 41 | 72   | 16.3 | 1068 | 5 Q8WRN3  | Q8WRN3 giardia lam |
| 42 | 72   | 16.3 | 1109 | 5 Q95V21  | Q95V21 giardia lam |
| 43 | 72   | 16.3 | 1423 | 6 Q7YS39  | Q7YS39 felis silve |
| 44 | 71.5 | 16.2 | 309  | 10 Q7XNY2 | Q7XNY2 oryza sativ |
| 45 | 71.5 | 16.2 | 626  | 4 Q8N1F3  | Q8N1F3 homo sapien |

## ALIGNMENTS

### RESULT 1

|        |  |        |              |      |         |
|--------|--|--------|--------------|------|---------|
| Q7YR42 | ID   | Q7YR42 | PRELIMINARY; | PRT; | 156 AA. |
| AC     | Q7YR42;  |        |              |      |         |
| DT     | 01-OCT-2003 (T-EMBLrel. 25, Created)                                   |        |              |      |         |
| DT     | 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)                      |        |              |      |         |
| DT     | 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)                    |        |              |      |         |
| DE     | IBX-1 protein.   |        |              |      |         |
| GN     | IBX-1.   |        |              |      |         |
| OS     | Pan troglodytes (Chimpanzee).  |        |              |      |         |
| OC     | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;      |        |              |      |         |
| OC     | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.              |        |              |      |         |
| OX     | NCBI_TaxID=9596;   |        |              |      |         |
| RN     | [1]  |        |              |      |         |
| RP     | SEQUENCE FROM N.A.   |        |              |      |         |
| FX     | MEDLINE=22709134; PubMed=12799463;                                     |        |              |      |         |
| RA     | Anzai T., Shihina T., Kimura N., Yanagiya K., Kohara S., Shigenari A., |        |              |      |         |
| RA     | Yamagata T., Kulski J.K., Naruse T.K., Fujimori Y., Fukuzumi Y.,       |        |              |      |         |
| RA     | Yamazaki M., Tashiro H., Iwamoto C., Umebara Y., Imanishi T.,          |        |              |      |         |
| RA     | Meyer A., Ikeo K., Gojobori T., Bahram S., Inoko H.                    |        |              |      |         |
| RT     | "Comparative sequencing of human and chimpanzee MHC class I regions    |        |              |      |         |
| RT     | unveils insertions/deletions as the major path to genomic              |        |              |      |         |
| RT     | divergence."   |        |              |      |         |
| RL     | Proc. Natl. Acad. Sci. U.S.A. 100:7708-7713(2003).                     |        |              |      |         |
| DR     | EMBL; AB100084; BAC78173.1; .  |        |              |      |         |
| SQ     | SEQUENCE 156 AA; 16929 MW; 83C067CDCAC09650 CRC64;                     |        |              |      |         |

Query Match 100.0%; Score 442; DB 6; Length 156;  
Best Local Similarity 100.0%; Pred. No. 2e-39;  
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

|    |     |  |
|----|-----|--|
| Qy | 1   | VRRQLPVEENPAKLLFLLTTVFQIIMAEHGVPAIPDPAPNAASLAPTPVSPVLE 60  |
| Db | 71  | VRRQLPVEENPAKLLFLLTTVFQIIMAEHGVPAIPDPAPNAASLAPTPVSPVLE 130 |
| Qy | 61  | PFNLTSEPSDYALDLSFLQQHPAAF 86                               |
| Db | 131 | PFNLTSEPSDYALDLSFLQQHPAAF 156                              |

```

RESULT 2
O75353 PRELIMINARY; PRT; 193 AA.
AC O75353;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 23, Last annotation update)
DE ANTI-death protein.
GN IEX-1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98369175; PubMed=9703517;
RA Wu X.X., Ao Z., Prasad K.V.S., Wu R., Schlossman S.F.;
RT "IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell
RT survival."
RL Science 281:1998-1001(1998).
DR EMBL; AF039067; AAC32558.1; -.
DR EMBL; AF071596; AAC72344.1; -.
DR GO; GO:0008189; Apoptosis inhibitor activity; NAS.
DR GO; GO:0006916; Pranti-apoptosis; NAS.
SQ SEQUENCE 193 AA; 21028 MW; 7927D9D3F7BC7C57 CRC64;

Query Match 100.0%; Score 442; DB 4; Length 193;
Best Local Similarity 100.0%; Pred. No. 2.5e-39;
Matches 86; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VRQLVPEPAPKRLFLILLIVFCQILMAEGVPAPLPDPDAPNAASLAPTVPVLE 60
DB 108 VRQLVPEPAPKRLFLILLIVFCQILMAEGVPAPLPDPDAPNAASLAPTVPVLE 167

QY 61 PNLITSEPSDYALDLSTFLQHPA 86
DB 168 PNLITSEPSDYALDLSTFLQHPA 193

QY 168 PNLITSEPSDYALDLSTFLQHPA 193
DB 168 PNLITSEPSDYALDLSTFLQHPA 193

RESULT 3
Q91VZ5 PRELIMINARY; PRT; 160 AA.
AC Q91VZ5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to immediate early response 3.
GN IER3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; BC006950; AAC06950.1; -.
DR EMBL; AK051003; BAC34493.1; -.
DR MGD; MGI:104914; Ier3.
SQ SEQUENCE 160 AA; 17655 MW; 0666DF96E75FCF4 CRC64;

Query Match 70.6%; Score 312; DB 11; Length 160;

```

```

Best Local Similarity 70.7%; Pred. No. 1.5e-25;
Matches 65; Conservative 5; Mismatches 14; Indels 8; Gaps 2;

QY 1 VRQLVPEPAPKRLFLILLIVFCQILMAEGVPAPLPDPDAPNAASLAPTVPV 56
DB 71 VRQLVPEPAPKRLFLILLIVFCQILMAEGVPAPLPDPDAPNAASLAPTVP 128
QY 57 --PVLEPFLNLTSEPSDYALDLSTFLQHPA 86
DB 129 APPVLEPFLNLTSEPSDYALDLKAFIQQHPA 160

RESULT 4
Q96CZ8 PRELIMINARY; PRT; 335 AA.
AC Q96CZ8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013612; AAH13612.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 335 AA; 37028 MW; 2295943B9B212862 CRC64;

Query Match 19.1%; Score 84.5; DB 4; Length 335;
Best Local Similarity 33.3%; Pred. No. 0.61;
Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;

QY 7 VEEPAPKRLFL----LLTVFCQILMAEGVP-----APLPDPDAPNAASLAP----- 52
DB 121 LQEHRLQKRLFLQKQQLQAYFNQMAESSYPQSQQLPLPROETPPSQPPPSLT 180

QY 53 TPUSPVLEPFLNLTSEPSDYALDLSTFLQHP 82
DB 181 QPLSPVLEP---SSQMQY----SPFLSQY 203

RESULT 5
Q9H0K1 PRELIMINARY; PRT; 926 AA.
AC Q9H0K1; O94878;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKP2p434K115.
GN DKP2P434K115.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansorge W., Boecker M., Bloeker H., Baurachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Vewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."
RL Genome Res. 11:422-435(2001).
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

```

```

DR EMEL; AL136764; CAB66698.1; -.
DR HSPF; Q63450; IAO6.
DR GO; GO:000524; F:protein binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser Thr kinase.
DR InterPro; IPR008271; Ser Thr pkin_AS.
DR InterPro; IPR001245; Tyr pkinase.
DR InterPro; IPR000449; UBA domain.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00030; UBA; 1.
KW Hypothetical protein; ATP-binding; Kinase;
KW Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 926 AA; 103915 MW; 9D7F9BF81FD65CCF CRC64;

Query Match 19.1%; Score 84.5; DB 4; Length 926;
Best Local Similarity 33.3%; Pred. No. 1.8;
Matches 30; Conservative 12; Mismatches 27; Indels 21; Gaps 5;

QY 7 VEENPAKRLFL-----LTIVFCQILMAEGVP-----APLPEDAPNAASLAP----- 52
DQ 712 LQEHRLQKQLFLQKQQLQAVFNQWQIAESSYQPSQQLPLPQETPPPSQAPPSLT 771
QY 53 TPVSVLPFPNLTSPSDYALDLSTFLQ 82
DQ 772 QPLSPVLEP-----SSEQMRY-----SPFLSQY 794

RESULT 6
Q9TZD3
ID Q9TZD3 PRELIMINARY; PRT; 102 AA.
AC Q9TZD3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN R160.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA "None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Nelson J., Gattung S.;
RT "The sequence of C. elegans cosmid R160.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMEL; AF090001; AAC68734.1; -.
DR PIR; T33565; T33565.
DR WormPep; R160.5; CE19555.

```

```

KW Hypothetical protein.
SQ SEQUENCE 102 AA; 11426 MW; AFD366C3932EB26C CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 132;
Best Local Similarity 24.7%; Pred. No. 0.36;
Matches 19; Conservative 19; Mismatches 28; Indels 11; Gaps 3;

QY 11 NPAKRLFL-----LTIVFCQILMAEGVPAPLPEDAPNAASLAPTVSPVLEPN-- 53
DQ 7 NPAIQAFVVLVCAVLCIEMERADGFCPLPPEELRIWNTIP-PTPIAAPVNR 65
QY 64 ---LTSEPSDYALDLST 77
DQ 66 QVWIKIDPTESVDWTS 82

RESULT 7
Q9UAS5
ID Q9UAS5 PRELIMINARY; PRT; 452 AA.
AC Q9UAS5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE Developmental-specific protein conZa8 (Fragment).
GN CONZA8.
OS Euplotes crassus.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OC Euplotida; Euplotidae; Moneuplotes.
OX NCBI_TaxID=5936;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ST9;
RX MEDLINE=9726572; PubMed=9172827;
RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;
RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene
expression from the old macronucleus.";
RL J. Eukaryot. Microbiol. 44:1-11(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ST9;
RX "conZa8 encodes a novel and abundant protein targeted to the
developing macronucleus in Euplotes crassus.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMEL; AF116196; RAD29625.1; -.
FT NON TER 1
SQ SEQUENCE 452 AA; 49763 MW; A5C13C49ECA998EA CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 452;
Best Local Similarity 34.5%; Pred. No. 1.7;
Matches 30; Conservative 10; Mismatches 30; Indels 17; Gaps 5;

QY 3 RQLPVEEP---NPAKRLFLLTIVFCQILMAEGVPAPLPEDAPNAASLAP-----P 54
DQ 316 RQSTPLPDILENPAKISAGSLSKMI---QETATESIPTQPP-----VSSIQSDAVHP 367
QY 55 VSPVLEPN-LTSEPSDYALDLSTFLQ 80
DQ 368 SAPSLPSTTASEPRDIIIPPSALK 394

RESULT 8
Q9XY65
ID Q9XY65 PRELIMINARY; PRT; 468 AA.
AC Q9XY65;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONZA8.
GN CONZA8.
OS Euplotes crassus.
OC Eukaryota; Alveolata; Ciliophora; Spirotrichea; Hypotrichia;
OC Euplotida; Euplotidae; Moneuplotes.

```

```

OX NCBI_TaxID=5936;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=G1;
RX MEDLINE=97206572; PubMed=9172827;
RA Ling Z., Ghosh S., Jacobs M.E., Klobutcher L.A.;
RT "Conjugation-specific genes in the ciliate Euplotes crassus: gene
expression from the old macronucleus.";
RL J. Eukaryot. Microbiol. 44:1-11(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=G1;
RX MEDLINE=98391748; PubMed=9722644;
RA Klobutcher L.A., Gyax S.E., Podoloff J.D., Vermeesch J.R.,
RT Price C.M., Tebeau C.M., Jahn C.L.;
RT "Conserved DNA sequences adjacent to chromosome fragmentation and
telomere addition sites in Euplotes crassus.";
RL Nucleic Acids Res. 26:4230-4240(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=G1;
RA Ling Z., Klobutcher L.A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF061334; AAD31724.1; -.
SQ SEQUENCE 468 AA; 51457 MW; 70996125317C592A CRC64;

Query Match 18.4%; Score 81.5; DB 5; Length 468;
Best Local Similarity 34.5%; Pred. No. 1.8; Mismatches 30; Indels 17; Gaps 5;
Matches 30; Conservative 10;

QY 3 RQLVDEP-----NPAKRLFLTLITVFCQILMAEGVPAPLPDPAPNAASLAPT-----P 54
DB 332 RQSPFLDILELPAKIASLSKMI---QETALESIPTQPP-----VSIQPSAPVHP 383

QY 55 VSPVLEPFPNLTSPSDYALDLSTFLQ 80
DB 384 SAPSPSPSTLASEPRDIIPTDPSAALK 410

RESULT 9
Q8CFH6 PRELIMINARY; PRT; 931 AA.
AC Q8CFH6;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Salt inducible kinase 2.
GN G630080D20RIK OR SIK2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adipose tissue;
RA Takemori H., Horike N., Katoh Y., Doi J., Lin X., Okamoto M.;
RT "Mouse salt inducible kinase 2 (SIK2).";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB067780; BAC53845.1; -.
DR MGD; MGI:2445031; SIK2.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0004658; P:protein amino acid phosphorylation; IEA.
DR InterPro; IPR000719; Prot kinase.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR001245; Tyr_kinase.
DR InterPro; IPR000449; UBA domain.
DR Pfam; PF000059; pkinase; I.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.

```

```

DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS50030; UBA; 1.
KW Kinase.
SQ SEQUENCE 931 AA; 104198 MW; 5CF2FB8DCG689F4 CRC64;

Query Match 18.4%; Score 81.5; DB 11; Length 931;
Best Local Similarity 29.7%; Pred. No. 3.7; Mismatches 29; Indels 21; Gaps 4;
Matches 27; Conservative 14;

QY 7 VERPNPAKELLFL---LTIIVFCQILMAEGVPAP-----LPPEDAPNAASLAPT----- 53
DB 712 LQHRLOQRFLQKQSQLOAYFNQMOIAESSYFPCSQQLALPHQETPLTSQQPFSFSLT 771
QY 54 -PVSPLYEPFNLTSPESDYALDLSTFLQOHP 83
DB 772 QALSPVL-----EPSSEQMCFSSFLSQYP 795

RESULT 10
Q9QM69 PRELIMINARY; PRT; 302 AA.
AC Q9QM69;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE 33 Kda protein.
OS Avian adenovirus type 8 (strain ATCC A-2A) (Fowl adenovirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Aviadenovirus.
OX NCBI_TaxID=66295;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20318671; PubMed=10859390;
RA Ojicic D., Nagy E.;
RT "The complete nucleotide sequence of fowl adenovirus type 8.";
RL J. Gen. Virol. 81:1833-1837(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Ojicic D., Nagy E.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF083975; AAD50348.1; -.
SQ SEQUENCE 302 AA; 34763 MW; BF34DF42FF026CCC CRC64;

Query Match 18.2%; Score 80.5; DB 12; Length 302;
Best Local Similarity 26.4%; Pred. No. 1.4; Mismatches 30; Indels 11; Gaps 4;
Matches 24; Conservative 26;

QY 2 RROQVVEEPNPARKLLFLTLITVFCQILMAEGVPAPLP-----PEDAPNAASLAPT 53
DB 171 RKLLDLTDSDDRRSPPFMF-LVSDDAVIDSEVPVPLPKRRRRANKKEPNASETOLPE 229

QY 54 PVSPLYEPFNLTSPESDYALDLSTFLQOHP 84
DB 230 PVSFAVS--DLKABTLNLLVEIESFVRKNPS 256

RESULT 11
Q8ITK2 PRELIMINARY; PRT; 343 AA.
AC Q8ITK2;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Son of sevenless (Fragment).
GN SOS OR CG793.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RA Mousset S., Brazier L., Cariou M.-L., Chartois F., Depaulis F.,
RA Veuille M.;
RT "Evidence of a high rate of selective sweeps in African Drosophila
RT melanogaster.";
RL Genetics 0:0-0(2002).
DR EMBL; AF459570; AAN61386.1; -.
DR FlyBase; FBgn0001965; Sos.
FT NON_TER 1 343
FT NON_TER 343 343
SQ SEQUENCE 343 AA; 36922 MW; CFS455BA3E9B779 CRC64;

Query Match 18.2%; Score 80.5; DB 5; Length 343;
Best Local Similarity 28.3%; Pred. No. 1.7; Mismatches 15; Indels 25; Gaps 6;
Matches 30; Conservative 15;

QY 1 VRQLPVE-----EPNPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPT 53
Db 184 IRNSAIEKRAAATSQPNQAAGPSTITLVTSQAVATDEPLPLISP-----AASSSTT 238

QY 54 --PVSVLPLEPP--NLTSEP-----SDYALDLSTFLQO----HPAAF 86
Db 239 TSLTPAMSPMSNPISHPVSTSSYAHQLRMQOQQOQTHPAYI 284

RESULT 12
Q8ITK1 PRELIMINARY; PRT; 343 AA.
AC Q8ITK1;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Son of sevenless (Fragment).
GN SOS.
OS Drosophila simulans (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7240;
RN [1]
RP SEQUENCE FROM N.A.
RA Mousset S., Brazier L., Cariou M.-L., Chartois F., Depaulis F.,
RA Veuille M.;
RT "Evidence of a high rate of selective sweeps in African Drosophila
RT melanogaster.";
RL Genetics 0:0-0(2002).
DR EMBL; AF459586; AAN61402.1; -.
DR FlyBase; FBgn0063991; Dsim\Sos.
FT NON_TER 1 343
FT NON_TER 343 343
SQ SEQUENCE 343 AA; 36909 MW; F80FFCDBE552FIAL CRC64;

Query Match 18.0%; Score 79.5; DB 5; Length 343;
Best Local Similarity 28.3%; Pred. No. 2.1;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRQLPVE-----EPNPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPT 53
Db 184 IRNSAIEKRAAATSQPNQAAGPSTITLVTSQAVATDEPLPLISP-----AASSSTT 238

QY 54 --PVSVLPLEPP--NLTSEP-----SDYALDLSTFLQO----HPAAF 86
Db 239 TSLTPAMSPMSNPISHPVSTSSYAHQLRMQOQQOQTHPAYI 284

RESULT 13
Q8IOF7 PRELIMINARY; PRT; 343 AA.
AC Q8IOF7;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Son of sevenless (Fragment).

```

```

GN SOS OR CG7793.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Mousset S., Brazier L., Cariou M.-L., Chartois F., Depaulis F.,
RA Veuille M.;
RT "Evidence of a high rate of selective sweeps in African Drosophila
RT melanogaster.";
RL Genetics 0:0-0(2002).
DR EMBL; AF459578; AAN61394.1; -.
DR EMBL; AF459579; AAN61395.1; -.
DR FlyBase; FBgn0001965; Sos.
FT NON_TER 1 343
FT NON_TER 343 343
SQ SEQUENCE 343 AA; 36895 MW; 85B12487BEF7310B CRC64;

Query Match 18.0%; Score 79.5; DB 5; Length 343;
Best Local Similarity 28.3%; Pred. No. 2.1;
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRQLPVE-----EPNPAKRLIFLLITVFCQILMAEGVPALPPEDAPNAASLAPT 53
Db 184 IRNSAIEKRAAATSQPNQAAGPSTITLVTSQAVATDEPLPLISP-----AASSSTT 238

QY 54 --PVSVLPLEPP--NLTSEP-----SDYALDLSTFLQO----HPAAF 86
Db 239 TSLTPAMSPMSNPISHPVSTSSYAHQLRMQOQQOQTHPAYI 284

RESULT 14
Q8IOG7 PRELIMINARY; PRT; 343 AA.
ID Q8IOG7;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TEMBLrel. 24, Last annotation update)
DE Son of sevenless (Fragment).
GN SOS OR CG7793.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Mousset S., Brazier L., Cariou M.-L., Chartois F., Depaulis F.,
RA Veuille M.;
RT "Evidence of a high rate of selective sweeps in African Drosophila
RT melanogaster.";
RL Genetics 0:0-0(2002).
DR EMBL; AF459566; AAN61382.1; -.
DR EMBL; AF459567; AAN61383.1; -.
DR EMBL; AF459568; AAN61384.1; -.
DR EMBL; AF459569; AAN61385.1; -.
DR EMBL; AF459571; AAN61387.1; -.
DR EMBL; AF459572; AAN61388.1; -.
DR EMBL; AF459573; AAN61389.1; -.
DR EMBL; AF459574; AAN61390.1; -.
DR EMBL; AF459575; AAN61391.1; -.
DR EMBL; AF459576; AAN61392.1; -.
DR EMBL; AF459577; AAN61393.1; -.
DR EMBL; AF459580; AAN61396.1; -.
DR EMBL; AF459581; AAN61397.1; -.
DR EMBL; AF459582; AAN61398.1; -.
DR EMBL; AF459583; AAN61399.1; -.
DR EMBL; AF459584; AAN61400.1; -.
DR EMBL; AF459585; AAN61401.1; -.
DR FlyBase; FBgn0001965; Sos.
FT NON_TER 1 343
FT NON_TER 343 343

```

FT NON TER 343 343  
SQ SEQUENCE 343 AA; 36879 MW; E86B80944FB85763 CRC64;

Query Match 18.0%; Score 79.5; DB 5; Length 343;  
Best Local Similarity 28.3%; Pred.No.2.1;  
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRQLPVE-----EPNPAKRLIFLLITIVFCQILMAERGVPAPLPEDAPNAASLAPT 53  
Db 184 IRRNSAIEKRAAATSQPNQAAGPISITLTVTSQAVATDEPLPLPISP-----AASSSTT 238  
QY 54 --PVSPVLEPP--NLTSEP-----SDYALDLSTFLQO-----HPAAP 86  
Db 239 TSPLTPAMSPWSPNIPSPHPVSTSSSYAHQLRMFQQQQOQTHPAIY 284

## RESULT 15

Q95TI8 PRELIMINARY; PRT; 372 AA.

AC Q95TI8;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE SD01193P.  
GN SOS OR CG793.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_FaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
RA Nuncio J., Paoleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,  
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY058749; AAL13978.1; .  
DR FlyBase; FBgn0001965; SOS.  
SQ SEQUENCE 372 AA; 40123 MW; D9CA3F33A2954F12 CRC64;

Query Match 18.0%; Score 79.5; DB 5; Length 372;  
Best Local Similarity 28.3%; Pred.No.2.3;  
Matches 30; Conservative 14; Mismatches 37; Indels 25; Gaps 6;

QY 1 VRQLPVE-----EPNPAKRLIFLLITIVFCQILMAERGVPAPLPEDAPNAASLAPT 53  
Db 199 IRRNSAIEKRAAATSQPNQAAGPISITLTVTSQAVATDEPLPLPISP-----AASSSTT 253  
QY 54 --PVSPVLEPP--NLTSEP-----SDYALDLSTFLQO-----HPAAP 86  
Db 254 TSPLTPAMSPWSPNIPSPHPVSTSSSYAHQLRMFQQQQOQTHPAIY 299

Search completed: February 25, 2004, 05:36:03  
Job time : 41 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 3, 2004, 17:21:43 ; Search time 1614 Seconds  
(without alignments)  
6928.440 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258  
Sequences: 1 GTCGGGGCCAGCTGCCAGT.....ACCAACACCCGGCGCCCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pla.\*
- 35: em.htg.rod.\*
- 36: em.htg.man.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID       | Description         |
|------------|-------|-------|--------|----|----------|---------------------|
| 1          | 258   | 100.0 | 471    | 9  | BT006703 | BT006703 Homo sapi  |
| 2          | 258   | 100.0 | 471    | 12 | BT007998 | BT007998 Synthetic  |
| 3          | 258   | 100.0 | 477    | 6  | AX392420 | AX392420 Sequence   |
| 4          | 258   | 100.0 | 477    | 9  | AF083421 | AF083421 Homo sapi  |
| 5          | 258   | 100.0 | 1228   | 6  | AB103436 | AB103436 Sequence   |
| 6          | 258   | 100.0 | 1228   | 6  | AR201462 | AR201462 Sequence   |
| 7          | 258   | 100.0 | 1228   | 6  | AR265155 | AR265155 Sequence   |
| 8          | 258   | 100.0 | 1228   | 6  | BD223034 | BD223034 Compositi  |
| 9          | 258   | 100.0 | 1230   | 9  | HSDIF2   | Y14551 Homo sapien  |
| 10         | 258   | 100.0 | 1238   | 9  | BC000844 | BC000844 Homo sapi  |
| 11         | 258   | 100.0 | 1240   | 9  | BC005080 | BC005080 Homo sapi  |
| 12         | 258   | 100.0 | 1309   | 9  | AF039067 | AF039067 Homo sapi  |
| 13         | 258   | 100.0 | 1693   | 9  | AF071596 | AF071596 Homo sapi  |
| 14         | 258   | 100.0 | 1864   | 9  | HSPRG1   | X96438 H.sapiens p  |
| 15         | 258   | 100.0 | 19459  | 9  | AB088101 | AB088101 Homo sapi  |
| 16         | 258   | 100.0 | 44118  | 9  | AC006165 | AC006165 Homo sapi  |
| 17         | 258   | 100.0 | 54911  | 9  | BC248307 | BC248307 Human DNA  |
| 18         | 258   | 100.0 | 90244  | 9  | AB023051 | AB023051 Homo sapi  |
| 19         | 258   | 100.0 | 130755 | 9  | AL845353 | AL845353 Human DNA  |
| 20         | 258   | 100.0 | 185617 | 9  | AL662848 | AL662848 Human DNA  |
| 21         | 258   | 100.0 | 200000 | 9  | AF000512 | AF000512 Homo sapi  |
| 22         | 258   | 100.0 | 300050 | 9  | AB100084 | AB100084 Pan trogl  |
| 23         | 256.4 | 99.4  | 1223   | 6  | AX410391 | AX410391 Sequence   |
| 24         | 256.4 | 99.4  | 1223   | 9  | S81914   | S81914 IEX-1-radia  |
| 25         | 256.4 | 99.4  | 171627 | 9  | AL662797 | AL662797 Human DNA  |
| 26         | 236   | 91.5  | 1232   | 6  | AX202085 | AX202085 Sequence   |
| 27         | 185.8 | 73.6  | 349980 | 6  | AX344553 | AX344553 Sequence   |
| 28         | 156   | 60.5  | 220314 | 2  | AC097188 | AC097188 Rattus no  |
| 29         | 156   | 60.5  | 297409 | 2  | AC096275 | AC096275 Rattus no  |
| 30         | 156   | 60.5  | 300988 | 2  | AC096249 | AC096249 Rattus no  |
| 31         | 154.4 | 59.8  | 1758   | 6  | AX827923 | AX827923 Sequence   |
| 32         | 154.4 | 59.8  | 1758   | 10 | RNPRG1   | X95437 R.horveglicu |
| 33         | 151.2 | 58.6  | 8188   | 10 | AY168443 | AY168443 Mus muscu  |
| 34         | 151.2 | 58.6  | 203476 | 2  | AC074150 | AC074150 Mus muscu  |
| 35         | 149.6 | 58.0  | 1110   | 10 | BC006950 | BC006950 Mus muscu  |
| 36         | 149.6 | 58.0  | 1938   | 10 | MNGLY96  | X67644 M.musculus   |
| 37         | 133.2 | 54.0  | 175345 | 2  | AC022301 | AC022301 Mus muscu  |
| 38         | 119.4 | 46.3  | 349980 | 6  | AX344571 | AX344571 Sequence   |
| 39         | 118   | 45.7  | 136981 | 10 | BX001064 | BX001064 Mouse DNA  |
| 40         | 118   | 45.7  | 278343 | 2  | EX571882 | EX571882 Mus muscu  |
| 41         | 110   | 42.6  | 174535 | 9  | AC106872 | AC106872 Homo sapi  |
| 42         | 110   | 42.6  | 191312 | 9  | AC023772 | AC023772 Homo sapi  |
| 43         | 84    | 32.6  | 193    | 6  | BD071420 | BD071420 Secreted   |
| 44         | 44    | 17.1  | 297    | 6  | A75448   | A75448 Sequence 11  |
| 45         | 44    | 17.1  | 297    | 6  | A78427   | A78427 Sequence 11  |

ALIGNMENTS

|            |   |             |      |        |                 |
|------------|---|-------------|------|--------|-----------------|
| RESULT 1   | BT006703  | 471 bp      | mRNA | linear | PRI 13-MAY-2003 |
| LOCUS      | Homo sapiens immediate early response 3 mRNA, complete cds.   |             |      |        |                 |
| DEFINITION | BT006703  |             |      |        |                 |
| ACCESSION  | BT006703.1  | GI:30582244 |      |        |                 |
| VERSION    | FLI_CDNA  |             |      |        |                 |
| KEYWORDS   | Homo sapiens (human)  |             |      |        |                 |
| SOURCE     | Homo sapiens  |             |      |        |                 |
| ORGANISM   | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  |             |      |        |                 |
| REFERENCE  | 1 (bases 1 to 471)  |             |      |        |                 |
| AUTHORS    | Kalnine.N., Chen.X., Rolfs.A., Halleck.A., Hines.L., Eisenstein.S., Koundirya.M., Raphael.J., Moreira.D., Kelley.T., LaBaer.J., Lin.Y., Phelan.M. and Farmer.A. |             |      |        |                 |



TITLE Cloning of human full-length CDSs in BD Creator (TM) System Donor vector

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 471)

AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA

COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion (TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orfclones>.

FEATURES

source

1..471

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="GH006131.0"

/clone\_lib="BD Creator (TM) CDS Library derived from MGC collection"

/lab\_host="DH5alpha TI resistant"

/note="Vector: pDNR-Dual"

CDS

1..471

/codon\_start=1

/product="immediate early response 3"

/protein\_id="AAP35349.1"

/db\_xref="GI:3058245"

/translations="MCHSRSGHPTWTLQAPTPAPSTIPGRSGPEFTFDPLPEP AAPGRPSASRGHRSRVLYPRVRLPVEPNPAKELLFLLLTIVFCQLMAE EGVPAFLPPEDAPNAAASLAPTPVSPVLEPFLNLTSPSDYALDLSTFLQHPAAF"

ORIGIN

Query Match 100.0%; Score 258; DB 9; Length 471;

Best Local Similarity 100.0%; Pred. No. 3.3e-49;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCT 120

DB 211 GTCCGGCGCCAGCTGCAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCT 270

QY 61 CTCACCATCGTCTTCTGCGAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCT 120

DB 271 CTCACCATCGTCTTCTGCGAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCT 330

QY 121 CCAGAGAGCGCCCTTAAGCGCGATCCCTGCGCGCCCTGCTGCTCCCTCGAG 180

DB 331 CCAGAGAGCGCCCTTAAGCGCGATCCCTGCGCGCCCTGCTGCTCCCTCGAG 390

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGACATTTCTCCAG 240

DB 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGACATTTCTCCAG 450

QY 241 CAACACCGCGCGCTTC 258

DB 451 CAACACCGCGCGCTTC 468

RESULT 2

BT007998

LOCUS BT007998 471 bp mRNA linear SYN 13-MAY-2003

DEFINITION Synthetic construct Homo sapiens immediate early response 3 mRNA, partial cds.

ACCESSION BT007998

VERSION BT007998.1 GI:30584834

KEYWORDS FLI\_CDNA.

SOURCE synthetic construct

ORGANISM synthetic construct

REFERENCE 1 (bases 1 to 471)

AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.

TITLE Cloning of human full-length CDSs in BD Creator (TM) System Donor vector

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 471)

AUTHORS Kalnine,N., Chen,X., Rolfs,A., Halleck,A., Hines,L., Eisenstein,S., Koundinya,M., Raphael,J., Moreira,D., Kelley,T., LaBaer,J., Lin,Y., Phelan,M. and Farmer,A.

TITLE Direct Submission

JOURNAL Submitted (13-MAY-2003) BD Biosciences Clontech, 1020 East Meadow Circle, Palo Alto, CA 94303, USA

COMMENT This CDS clone is a part of a collection of human full length expression clones generated by BD Biosciences Clontech and the Harvard Institute of Proteomics. Each CDS has been cloned in two forms: with and without stop-codon (to allow fusion with C-terminal tag). The CDS has been directionally cloned using BD In-Fusion (TM) cloning system between the SalI and HindIII sites of the pDNR-DUAL vector. Additional sequences in the clone: 'ACC' after SalI site and before 'ATG' to provide Kozak consensus sequence; 'GG' after last codon and before HindIII site to maintain reading frame. Clone distribution: <http://bioinfo.clontech.com/orfclones>.

FEATURES

source

1..471

/organism="synthetic construct"

/mol\_type="mRNA"

/db\_xref="taxon:32630"

/clone="GH006131.0"

/clone\_lib="BD Creator (TM) CDS Library derived from MGC collection"

/lab\_host="DH5alpha TI resistant"

/note="Vector: pDNR-Dual"

CDS

1..471

/note="Mutations: 470:Stop->Leu"

/codon\_start=1

/transl\_table=11

/product="Homo sapiens immediate early response 3"

/protein\_id="AAP36670.1"

/db\_xref="GI:30584835"

/translations="MCHSRSGHPTWTLQAPTPAPSTIPGRSGPEFTFDPLPEP AAPGRPSASRGHRSRVLYPRVRLPVEPNPAKELLFLLLTIVFCQLMAE EGVPAFLPPEDAPNAAASLAPTPVSPVLEPFLNLTSPSDYALDLSTFLQHPAAF"

ORIGIN

Query Match 100.0%; Score 258; DB 12; Length 471;

Best Local Similarity 100.0%; Pred. No. 3.3e-49;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTGCAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCTG 60

DB 211 GTCCGGCGCCAGCTGCAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCTG 270

QY 61 CTCACCATCGTCTTCTGCGAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCT 120

DB 271 CTCACCATCGTCTTCTGCGAGATCCCTGATGCTGAAGGGGTGCGCGCCCTGCCT 330

QY 121 CCAGAGAGCGCCCTTAAGCGCGATCCCTGCGCGCCCTGCTGCTCCCTCGAG 180

DB 331 CCAGAGAGCGCCCTTAAGCGCGATCCCTGCGCGCCCTGCTGCTCCCTCGAG 390

QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGACATTTCTCCAG 240

DB 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGACATTTCTCCAG 450

QY 241 CAACACCGCGCGCTTC 258

DB 451 CAACACCGCGCGCTTC 468

```

RESULT 3
AX392420
LOCUS AX392420 477 bp DNA linear PAT 23-MAR-2002
DEFINITION Sequence 4 from Patent WO0216416.
ACCESSION AX392420
VERSION AX392420.1 GI:19700736
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Lee,R.T., Landschulz,K.T., Kennedy,S.P., Thompson,J.F. and Turi,T.G.
Diagnosis and treatment of cardiovascular conditions
Patent: WO 0216416-A 4 28-FEB-2002;
THE BRIGHAM AND WOMEN'S HOSPITAL, INC. (US) ; PFIZER INC. (US)
FEATURES
source
1..477
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
7..477
/note="unnamed protein product; IEX1"
/codon_start=1
/protein_id="CAD29007.1"
/db_xref="GI:19700737"
/db_xref="PEMIREMBL:CAD29007"
/translation="MCHSRSCHEFTMTILQAPTPAPSTIPGPRGSGEFTFDPLPEP
AAPAGRSASGRHRKSRVLYPRVRRQLPVEPNPAKRLILFLITIVFOILMAE
EGVPAPLPEDAPNAASLAPTPVSPVLEPNLTSEPSDYALDLSTFLQHPAAF"
ORIGIN
Query Match 100.0%; Score 258; DB 6; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.3e-49;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCGGCGCCAGTCCAGTCGAGGAAACCGAACCCAGCAAGAGGCTTCTTTCTGCTG 60
Db 217 GTCCGGCGCCAGTCCAGTCGAGGAAACCGAACCCAGCAAGAGGCTTCTTTCTGCTG 276
QY 61 CTCACCATCGTCTTCTGCGAGATCTGATGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 120
Db 277 CTCACCATCGTCTTCTGCGAGATCTGATGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 336
QY 121 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180
Db 337 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 396
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGACCTTTCCTCCAG 240
Db 397 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGACCTTTCCTCCAG 456
QY 241 CAACACCGCGCGCTTC 258
Db 457 CAACACCGCGCGCTTC 474

RESULT 4
AF083421
LOCUS AF083421 477 bp mRNA linear PRI 12-OCT-2001
DEFINITION Homo sapiens radiation-inducible immediate early response gene IEX1
(IEX1) mRNA, complete cds.
ACCESSION AF083421
VERSION AF083421.1 GI:3511288
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Kondratyev,A.D., Chung,K.N. and Jung,M.O.

```

```

TITLE Identification and characterization of a radiation-inducible
Glycosylated human early-response gene
Cancer Res. 56 (7), 1498-1502 (1996)
JOURNAL MEDLINE
PUBMED 8603392
REFERENCE 2 (bases 1 to 477)
AUTHORS Kumar,R., Kobayashi,T., Warner,G.M., Wu,Y., Salisbury,J.L.,
Lingle,W. and Pittelkow,M.R.
TITLE A novel immediate early response gene, IEX-1, is induced by
ultraviolet radiation in human keratinocytes
Biochem. Biophys. Res. Commun. 253 (2), 336-341 (1998)
JOURNAL MEDLINE
PUBMED 99097237
REFERENCE 3 (bases 1 to 477)
AUTHORS Kumar,R., Pittelkow,M.R. and Warner,G.M.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1998) Nephrology Research, Mayo Clinic, 200 1st
Street SW, Rochester, MN 55905, USA
FEATURES
Location/Qualifiers
source
1..477
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="keratinocyte; fetal osteoblast"
1..477
/gene="IEX1"
7..477
/gene="IEX1"
/note="UV irradiation induced gene in keratinocytes"
/codon_start=1
/product="radiation-inducible immediate early response
gene IEX1"
/protein_id="AAG33793.1"
/db_xref="GI:3511289"
/translation="MCHSRSCHEFTMTILQAPTPAPSTIPGPRGSGEFTFDPLPEP
AAPAGRSASGRHRKSRVLYPRVRRQLPVEPNPAKRLILFLITIVFOILMAE
EGVPAPLPEDAPNAASLAPTPVSPVLEPNLTSEPSDYALDLSTFLQHPAAF"
167
/gene="IEX1"
/citation=[1]
/replace="g"
323
/gene="IEX1"
/citation=[1]
/replace="g"
conflict
conflict
ORIGIN
Query Match 100.0%; Score 258; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 3.3e-49;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTCCGGCGCCAGTCCAGTCGAGGAAACCGAACCCAGCAAGAGGCTTCTTTCTGCTG 60
Db 217 GTCCGGCGCCAGTCCAGTCGAGGAAACCGAACCCAGCAAGAGGCTTCTTTCTGCTG 276
QY 61 CTCACCATCGTCTTCTGCGAGATCTGATGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 120
Db 277 CTCACCATCGTCTTCTGCGAGATCTGATGCTGAAGAGGGTGTGCGCGCGCCCTGCCT 336
QY 121 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 180
Db 337 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCTCCGCTCGAG 396
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGACCTTTCCTCCAG 240
Db 397 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGACCTTTCCTCCAG 456
QY 241 CAACACCGCGCGCTTC 258
Db 457 CAACACCGCGCGCTTC 474
RESULT 5

```







mRNA (cDNA clone MGC:13037 IMAGE:3617088), complete cds.

ACCESSION BC005080  
VERSION BC005080.1 GI:13477228  
KEYWORDS MGC.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 1240)  
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shellen,C.M., Schuler,G.D., Altshuler,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Eshat,N.K., Hopkins,R.F., Jordan,K., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.W., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.P., Casavant,T.L., Schectz,T.B., Brownstein,M.J., Ustin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Iqbal,M.A., Peters,G.J., Abramson,R.D., Mullikin,J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Feherty,J., Helton,E., Kettelman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smalish,D.E., Schnerch,A., Schein,J.E., Jones,S.J., and Marra,M.A.  
Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)  
12477932  
2. (bases 1 to 1240)  
Strausberg,R.  
Direct Submission  
Submitted (26-MAR-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
NH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: ATCC/DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILML)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcsc.bc.ca](mailto:info@bcsc.bc.ca)  
Steven Jones, Jennifer Amano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letitia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saedi, Jacqueline Schein, Duane Smalish, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILML at: <http://image.llnl.gov>  
Series: IRAL Plate: 18 Row: g Column: 17  
This clone was selected for full length sequencing because it passed the following selection criteria: GenomesScan gene prediction.  
FEATURES  
source  
1. .1240  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="MGC:13037 IMAGE:3617088"  
/tissue\_type="Skin, melanotic melanoma."  
/clone\_lib="NIH MGC\_20"  
/lab\_host="DH10B-R"  
/note="Vector: pOT7"

gene  
1. .1240  
/gene="IER3"  
/note="synonyms: IEX1, DIF2, PRG1, GLY96, IEX-1L, IEX-1, DIF-2"  
/db\_xref="LocusID:8870"  
/db\_xref="WIM:602996"  
19..489  
/codon\_start=1  
/product="immediate early response 3, isoform short"  
/protein\_id="AAH05080.1"  
/db\_xref="GI:13477228"  
/db\_xref="LocusID:8870"  
/translators="WCHSSSCHPTWTLQAPTAPSTIPGRGSGPEITFDLPBP  
AAAPGRSASRGHRGSRVLYPRVRRQLPVEENPAKLLILLITIVFCQILMAE  
EGVPAPLPEDAPNAASLAPTFVSPVLEPFLNLTSEPSDYALDLSTFLQQHPAAF"

CDS  
Query Match 100.0%; Score 258; DB 9; Length 1240;  
Best Local Similarity 100.0%; Pred. No. 3.2e-49;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cyt 1 GTCGGCGGCAGTCGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Dbs 229 GTCGGCGGCAGTCGCCAGTCGAGGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 288  
Cyt 61 CTCACCATGCTCTTCTGCGAGTCCTGATGCTGAGGAGGCTGTCGGCGGCCTGCTT 120  
Dbs 289 CTCACCATGCTCTTCTGCGAGTCCTGATGCTGAGGAGGCTGTCGGCGGCCTGCTT 348  
Cyt 121 CCAGAGGAGCGCCCTTAACGCGCATCTCTGCGCGCCACCCCTGCTGTCCTGCTCGAG 180  
Dbs 349 CCAGAGGAGCGCCCTTAACGCGCATCTCTGCGCGCCACCCCTGCTGTCCTGCTCGAG 408  
Cyt 181 CCCTTAATCTGACTCGAGGCGCTCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 240  
Dbs 409 CCCTTAATCTGACTCGAGGCGCTCGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 468  
Cyt 241 CAACACCGCGCGCTTC 258  
Dbs 469 CAACACCGCGCGCTTC 486

RESULT 12  
AF039067  
LOCUS Homo sapiens anti-death protein (IEX-1L) mRNA, complete cds.  
DEFINITION AF039067  
ACCESSION AF039067  
VERSION AF039067.1 GI:3449375  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1309)  
AUTHORS Wu,M.X., Ao,Z., Prasad,K.V., Wu,R. and Schlossman,S.F.  
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival  
JOURNAL Science 281 (5379), 998-1001 (1998)  
MEDLINE 98369175  
PUBMED 9703517  
REFERENCE 2 (bases 1 to 1309)  
AUTHORS Ao,Z. and Wu,M.X.  
TITLE Direct Submission  
JOURNAL Submitted (17-DEC-1997) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA  
FEATURES  
source  
1. .1309  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
1. .1309  
/gene="IEX-1L"  
1. .582

```
/gene="IEX-1L"
/note="type II membrane protein"
/codon_start=1
/product="anti-death protein"
/protein_id="AAC32558.1"
/db_xref="GI:3449376"
/translators="MCHSRSCHPTMTILOAPTAPSTIPGRRGSGPEIFTFDPLPEP
AAAPAGPSASRGRKRRVSLPSSGEYRSGHSCALPWSHGTTRLOSLDITCLL
FCLPLVRQUPVEENPAKRLIFLLITVFCQILMAEGVPAPLPEDAPNAASLAFT
PVSPVLPFFNLITSEPSYALDLSLTFLOQHPAAF"
```

## ORIGIN

```
Query Match 100.0%; Score 258; DB 9; Length 1309;
Best Local Similarity 100.0%; Pred. No. 3.2e-49;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCCTTTCTGCTG 60
DB 322 GTCCGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCCTTTCTGCTG 381
QY 61 CTCACCATGCTTCTTCCAGATCCCTGATGCTGAAGAGGGTGTGCCGGCCCTTGCCT 120
DB 382 CTCACCATGCTTCTTCCAGATCCCTGATGCTGAAGAGGGTGTGCCGGCCCTTGCCT 441
QY 121 CCAGAGAGCGCCCTACCGCCGATCCCTGCGGCCACCCCTGTGTCCCGGCTCTCCGAG 180
DB 442 CCAGAGAGCGCCCTACCGCCGATCCCTGCGGCCACCCCTGTGTCCCGGCTCTCCGAG 501
QY 181 CCCTTTAATCTGACTTCGGAGCCCTGGGACTACGCTCTGGACCTCAGCACTTCCGCCAG 240
DB 502 CCCTTTAATCTGACTTCGGAGCCCTGGGACTACGCTCTGGACCTCAGCACTTCCGCCAG 561
QY 241 CAACACCGCGCGCTTC 258
DB 562 CAACACCGCGCGCTTC 579
```

## RESULT 13

```
AF071596
LOCUS AF071596 Homo sapiens apoptosis inhibitor (IEX-1L) gene, complete cds.
DEFINITION AF071596
ACCESSION AF071596
VERSION AF071596.1 GI:3851531
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1693)
AUTHORS Wu, M.X., Ao, Z., Prasad, K.V., Wu, R. and Schlossman, S.P.
TITLE IEX-1L, an apoptosis inhibitor involved in NF-kappaB-mediated cell survival
JOURNAL Science 281 (5379), 998-1001 (1998)
MEDLINE 98369175
PUBMED 9703517
REFERENCE 2 (bases 1 to 1693)
AUTHORS Wu, M.X. and Ao, Z.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1998) Tumor Immunology, Dana Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA
FEATURES
Source 1..1693
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
359..1693
/gene="IEX-1L"
359..1693
/gene="IEX-1L"
/product="apoptosis inhibitor"
385..966
/gene="IEX-1L"
/note="immediate early response protein"
```

```
/codon_start=1
/product="apoptosis inhibitor"
/protein_id="AAC72344.1"
/db_xref="GI:3851532"
/translators="MCHSRSCHPTMTILOAPTAPSTIPGRRGSGPEIFTFDPLPEP
AAAPAGPSASRGRKRRVSLPSSGEYRSGHSCALPWSHGTTRLOSLDITCLL
FCLPLVRQUPVEENPAKRLIFLLITVFCQILMAEGVPAPLPEDAPNAASLAFT
PVSPVLPFFNLITSEPSYALDLSLTFLOQHPAAF"
```

## ORIGIN

```
Query Match 100.0%; Score 258; DB 9; Length 1693;
Best Local Similarity 100.0%; Pred. No. 3.1e-49;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCCTTTCTGCTG 60
DB 706 GTCCGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCCTTTCTGCTG 765
QY 61 CTCACCATGCTTCTTCCAGATCCCTGATGCTGAAGAGGGTGTGCCGGCCCTTGCCT 120
DB 766 CTCACCATGCTTCTTCCAGATCCCTGATGCTGAAGAGGGTGTGCCGGCCCTTGCCT 825
QY 121 CCAGAGAGCGCCCTACCGCCGATCCCTGCGGCCACCCCTGTGTCCCGGCTCTCCGAG 180
DB 826 CCAGAGAGCGCCCTACCGCCGATCCCTGCGGCCACCCCTGTGTCCCGGCTCTCCGAG 885
QY 181 CCCTTTAATCTGACTTCGGAGCCCTGGGACTACGCTCTGGACCTCAGCACTTCCGCCAG 240
DB 886 CCCTTTAATCTGACTTCGGAGCCCTGGGACTACGCTCTGGACCTCAGCACTTCCGCCAG 945
QY 241 CAACACCGCGCGCTTC 258
DB 946 CAACACCGCGCGCTTC 963
```

## RESULT 14

```
HSP9G1
LOCUS HSP9G1 1864 bp DNA linear PRI 06-NOV-1998
DEFINITION H.sapiens PRG1 gene.
ACCESSION X96438
VERSION X96438.1 GI:2440072
KEYWORDS PRG1 gene.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Schafer, H., Diebel, J., Ault, A., Trauzold, A. and Schmidt, W.E.
TITLE The promoter of human p22/PACAP response gene 1 (PRG1) contains functional binding sites for the p53 tumor suppressor and for NF-kappaB
JOURNAL FEBS Lett. 436 (2), 139-143 (1998)
MEDLINE 98452926
PUBMED 9781666
REFERENCE 2 (bases 1 to 508)
AUTHORS Schafer, H., Trauzold, A., Lettau, P., Kalthoff, H., Foelsch, U.R. and Schmidt, W.E.
TITLE cDNA cloning and sequencing of a novel human early response gene and characterization of its expression in pancreatic carcinoma cells
JOURNAL Gastroenterology
REFERENCE 3
AUTHORS Trauzold, A.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1996) A. Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG
REMARK 4 (bases 1 to 1864)
REFERENCE Schafer, H.
AUTHORS Direct Submission
TITLE Submitted (23-SEP-1997) H.Schaefer, Trauzold, Laboratory of Molecular Gastroenterology, 1st Dept. of Medicine, University of
```

Kiel, Schittenhelmstrasse 12, Kiel, D-24105, FRG  
COMMENT On Sep 27, 1997 this sequence version replaced gi:1515291.  
FEATURES  
source  
1..1864  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/cell\_line="818-4"  
gene  
1..1564  
/gene="PRGI"  
Promoter  
1..562  
/gene="PRGI"  
mRNA  
join(563..801,914..1864)  
exon  
563..801  
/gene="PRGI"  
CDS  
join(592..801,914..1174)  
/gene="PRGI"  
/codon\_start=1  
/protein\_id="CA465304.1"  
/db\_xref="GI:2440073"  
/db\_xref="GOA:P46695"  
/translation="MCHSRSCHPMTILQAPTAPSTIPGPRGSGPEITFTDPLPEP  
AAAPAGPSASGRHRSRVLPVVRQLPVEEPNPAKELLFLITIVFCQILMAE  
EGVPAPLPEDAPNAASLAPFTVSPVLEPNLTSPSDYALDLSTFLQHPAAF"  
intron  
802..913  
/gene="PRGI"  
exon  
914..1864  
/gene="PRGI"  
/number=2  
ORIGIN  
Query Match 100.0%; Score 258; DB 9; Length 1864;  
Best Local Similarity 100.0%; Pred. No. 3,1e-49;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCGCTG 60  
Db 914 GTCCGGCGCCAGCTGCCAGTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCGCTG 973  
QY 61 CTCACCATGCTTCTTCCAGATCTTGATGCTGAAGAGGTGTGCGGGGCCCGCTGCT 120  
Db 974 CTCACCATGCTTCTTCCAGATCTTGATGCTGAAGAGGTGTGCGGGGCCCGCTGCT 1033  
QY 121 CCAGAGAGCGCCCTTAACCGCCGATCCCTGCGGCCCGCCACCCCTGTGTCCCGCTCGAG 180  
Db 1034 CCAGAGAGCGCCCTTAACCGCCGATCCCTGCGGCCCGCCACCCCTGTGTCCCGCTCGAG 1093  
QY 181 CCCTTTAATGACTCTGGAGCCCTCGGACTACGCTCTGACCTCAGACTTTCCTCCAG 240  
Db 1094 CCCTTTAATGACTCTGGAGCCCTCGGACTACGCTCTGACCTCAGACTTTCCTCCAG 1153  
QY 241 CAACACCGGCGCGCTTC 258  
Db 1154 CAACACCGGCGCGCTTC 1171  
RESULT 15  
AB088101  
LOCUS  
DEFINITION Homo sapiens PRGI, FLOT1 genes for immediate early response 3,  
flotillin 1, complete cds.  
ACCESSION AB088101  
VERSION AB088101.1 GI:27544397  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1  
Shiina,T., Ota,M., Katsuyama,Y., Hashimoto,N. and Inoko,H.  
Genome diversity in HLA: A new strategy for detection of genetic  
polymorphisms in expressed genes within the HLA class III and class  
I regions  
Unpublished  
2 (bases 1 to 19459)  
Shiina,T.  
Direct Submission  
Submitted (08-JUL-2002) Takashi Shiina, Tokai University School of  
Medicine, Molecular Life Science 2; Bohseidai, Isehara, Kanagawa  
259-1193, Japan [E-mail:tshiina@is.foc.u-tokai.ac.jp,  
Tel:81-463-93-1121, Fax:81-463-94-8884]  
Location/Qualifiers  
1..19459  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/map="6p21.31"  
/cell\_line="LKT3"  
/cell\_type="B cell"  
1797..2035  
/number=1  
exon  
join(1826..2035,2148..2408)  
/gene="PRGI"  
gene  
join(1826..2035,2148..2408)  
CDS  
/gene="PRGI"  
/note="expressed in pancreatic carcinoma"  
/codon\_start=1  
/product="immediate early response 3"  
/protein\_id="BAC54933.1"  
/db\_xref="GI:27544398"  
/translation="MCHSRSCHPMTILQAPTAPSTIPGPRGSGPEITFTDPLPEP  
AAAPAGPSASGRHRSRVLPVVRQLPVEEPNPAKELLFLITIVFCQILMAE  
EGVPAPLPEDAPNAASLAPFTVSPVLEPNLTSPSDYALDLSTFLQHPAAF"  
exon  
2148..3101  
/number=2  
exon  
3102..3817  
/number=1  
exon  
441..4197  
/number=2  
gene  
join(4155..4197,4477..4552,4640..4730,5011..5154,  
5546..5665,5753..5848,6034..6186,15242..15423,  
15615..15660,15777..15914,16156..16320,18197..18226)  
/gene="FLOT1"  
CDS  
join(4155..4197,4477..4552,4640..4730,5011..5154,  
5546..5665,5753..5848,6034..6186,15242..15423,  
15615..15660,15777..15914,16156..16320,18197..18226)  
/gene="FLOT1"  
/note="integral membrane component of caveolae"  
/codon\_start=1  
/product="Flotillin 1"  
/protein\_id="BAC54934.1"  
/db\_xref="GI:27544399"  
/translation="MEFTGCPNEMVYSGFRSPVPVAGRVFVLPCTIQIQRISLN  
TLTAVKSEKVTYTHGVPIISVTGLAOKVIOGNKMLAAACCMELGKTEAHTALE  
TLEGHQRALMAHMTVEEIKYQKQSFQSVKVASDLVNMGISVSYTLXDIHQDY  
LHSJGKARTAQVDARIGEAEKRDAGIREAKQKQKQSVQYLSLEIMAKAQDYEL  
KKAAYDIENVNTRAQADLAYQLQVAKTQQIIEQVQVQVVERAQVAVQVQRIARE  
KELEARVPKAEAEKYLERLAEAKSQLMQAEAEASVMRGEAEAFIAGARAE  
AEQMAKAEAPQLYQEAQOLDMLLEKLPQVABEISGLTSANKITLVSSGSGTWSAAK  
VNGEVDLILPELPEVERLTGVISQVNHKPLRTA"  
4477..4552  
/gene="FLOT1"  
exon  
/number=3  
4640..4730  
/gene="FLOT1"  
exon  
/number=4  
5011..5154  
/gene="FLOT1"  
exon  
/number=5  
5546..5665  
/gene="FLOT1"



```

exon      /number=6
5753..5848
/gene="FLOT1"
number=7
6034..6186
/gene="FLOT1"
number=8
15242..15423
/gene="FLOT1"
number=9
15615..15660
/gene="FLOT1"
number=10
15777..15914
/gene="FLOT1"
number=11
16156..16320
/gene="FLOT1"
number=12
18197..18613
number=13

```

ORIGIN

```

Query Match      100.0%; Score 258; DB 9; Length 19459;
Best Local Similarity 100.0%; Pred. No. 2.8e-49;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1  GTCCGGGCGCAGTCCAGTCGAGGAACCGAACCAGGCAAAAGGCTTCTTTCTGCTG 60
Db      2148 GTCCGGGCGCAGTCCAGTCGAGGAACCGAACCAGGCAAAAGGCTTCTTTCTGCTG 2207

Qy      61  CTCACCATGCTCTTGTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 120
Db      2208 CTCACCATGCTCTTGTGCCAGATCCTGATGGCTGAAGAGGCTGTGCCGGCGCCCTGCCT 2267

Qy      121 CCAGAGGAGCGCCCTTAACGGCGCATCCCTGGCGGCCACCCCTGTGTGCCCGCTCTCGAG 180
Db      2268 CCAGAGGAGCGCCCTTAACGGCGCATCCCTGGCGGCCACCCCTGTGTGCCCGCTCTCGAG 2327

Qy      181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTGAGACTTTCTCCAG 240
Db      2328 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTGAGACTTTCTCCAG 2387

Qy      241 CAACACCCGCGCGCTTC 258
Db      2388 CAACACCCGCGCGCTTC 2405

```

Search completed: March 3, 2004, 19:39:14  
Job time : 1622 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 3, 2004, 16:07:27 ; Search time 254 Seconds  
(without alignments)  
4315.099 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCCGCGCCGCTGCCAGT.....AGCACACCCGCGCCCTTC 258

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373563 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 5747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq 29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001as:\*

5: Geneseq2001bs:\*

6: Geneseq2002s:\*

7: Geneseq2003as:\*

8: Geneseq2003bs:\*

9: Geneseq2003cs:\*

10: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description        |
|------------|-------|-------------|--------|----|--------------------|
| 1          | 258   | 100.0       | 477    | 6  | ABK12138 Human CDN |
| 2          | 258   | 100.0       | 1228   | 2  | AAT94471 Human Fch |
| 3          | 258   | 100.0       | 1228   | 3  | AAZ50711 Nucleotid |
| 4          | 258   | 100.0       | 1297   | 4  | AAR34278 Human col |
| 5          | 258   | 100.0       | 1308   | 4  | AAO08380 Human sec |
| 6          | 258   | 100.0       | 1578   | 4  | AAK73551 Human imm |
| 7          | 258   | 100.0       | 1578   | 4  | AAK73553 Human imm |
| 8          | 258   | 100.0       | 1578   | 4  | AAK37434 Human mus |
| 9          | 258   | 100.0       | 1578   | 4  | AAK37436 Human mus |
| 10         | 258   | 100.0       | 1578   | 7  | ABX60422 cDNA enco |
| 11         | 258   | 100.0       | 1578   | 7  | ABX60424 cDNA enco |
| 12         | 256.4 | 99.4        | 1223   | 6  | ABK84167 Human cDN |
| 13         | 256.4 | 99.4        | 1223   | 6  | ABK64673 Human ben |
| 14         | 256.4 | 99.4        | 1223   | 6  | ABN96540 Human     |
| 15         | 165.4 | 64.1        | 251    | 4  | AAH35357 Gene #303 |
| 16         | 154.4 | 59.8        | 1758   | 9  | ABT42493 Toxicity  |
| 17         | 154.4 | 59.8        | 1758   | 9  | ABT42493 Toxicity  |
| 18         | 154.4 | 59.8        | 1758   | 9  | ABT42493 Toxicity  |
| 19         | 107.2 | 41.6        | 436    | 3  | AAK41806 Human sec |
| 20         | 103.6 | 40.2        | 324    | 7  | ABZ91442 Human nuc |
| 21         | 84    | 32.6        | 193    | 2  | AAV89271 EST clone |
| 22         | 60    | 23.3        | 60     | 6  | ABN32211 Human spl |
| 23         | 52.6  | 20.4        | 375    | 7  | ABX36077 Bovine ES |

|    |      |      |        |   |          |                    |
|----|------|------|--------|---|----------|--------------------|
| 24 | 44   | 17.1 | 297    | 2 | AAQ77534 | Aaq77534 Human gen |
| 25 | 39.6 | 15.3 | 9210   | 4 | AAH52046 | Aah52046 Mycobacte |
| 26 | 39.6 | 15.3 | 110000 | 4 | AAI99682 | Continuation (29 o |
| 27 | 39.6 | 15.3 | 110000 | 4 | AAI99683 | Continuation (29 o |
| 28 | 39.2 | 15.2 | 619    | 6 | ABQ53936 | Abq53936 Oligonuc  |
| 29 | 39.2 | 15.2 | 619    | 6 | ABQ53937 | Abq53937 Oligonuc  |
| 30 | 39.2 | 15.2 | 619    | 6 | ABQ53937 | Abq53937 Oligonuc  |
| 31 | 39.2 | 15.2 | 637    | 7 | ABT41296 | Abt41296 Toxicity  |
| 32 | 39.2 | 15.2 | 637    | 9 | ABD56978 | Abd56978 Toxicity  |
| 33 | 39.2 | 15.2 | 1038   | 3 | AAO7597  | Aao7597 Human cyc  |
| 34 | 39.2 | 15.2 | 1053   | 3 | AAO7595  | Aao7595 Human cyc  |
| 35 | 39.2 | 15.2 | 1092   | 3 | AAO7596  | Aao7596 Human cyc  |
| 36 | 39.2 | 15.2 | 2168   | 4 | AAI18833 | Aai18833 Human kin |
| 37 | 39.2 | 15.2 | 113193 | 7 | AAV54645 | Aav54645 Streptomy |
| 38 | 38.6 | 15.0 | 6354   | 2 | AAV69699 | Aav69699 5' UTR se |
| 39 | 38.6 | 15.0 | 6354   | 3 | AAV5724  | Aav5724 Human NEM  |
| 40 | 38.6 | 15.0 | 6354   | 6 | ABX53902 | Abx53902 Human iEN |
| 41 | 38.2 | 14.8 | 4141   | 5 | AAV78580 | Aav78580 DNA encod |
| 42 | 38.2 | 14.8 | 4228   | 3 | AAZ51683 | Aaz51683 Human cyc |
| 43 | 38.2 | 14.8 | 4340   | 6 | AAV94890 | Aav94890 Human DNA |
| 44 | 38.2 | 14.8 | 8391   | 7 | ACA64735 | Aca64735 Mycobacte |
| 45 | 37.6 | 14.6 | 170    | 2 | AAV69706 | Aav69706 Nucleotid |

ALIGNMENTS

|          |  |                         |
|----------|--|-------------------------|
| RESULT 1 | ABK12138   | standard; cDNA; 477 BP. |
| ID       | ABK12138   | standard; cDNA; 477 BP. |
| XX       | ABK12138   |                         |
| AC       | ABK12138   |                         |
| XX       | ABK12138   |                         |
| DT       | 05-JUN-2002  | (first entry)           |
| XX       | Human cDNA encoding proliferation regulated gene-1 protein, prg-1/IEX-1. |                         |
| DE       | Human cDNA encoding proliferation regulated gene-1 protein, prg-1/IEX-1. |                         |
| XX       | Human; ss; gene; MIVR-1; Mechanically Induced Vascular Receptor 1;       |                         |
| KW       | Cytostatic; cardiant; cerebroprotective; antiarteriosclerotic;           |                         |
| KW       | cardiac cell; anti-apoptotic; vascular endothelial cell;                 |                         |
| KW       | cardiac hypertrophy; myocardial infarction; stroke; arteriosclerosis;    |                         |
| KW       | heart failure; proliferation regulated gene-1; prg-1; IEX-1.             |                         |
| XX       | Homo sapiens.  |                         |
| XX       | Homo sapiens.  |                         |
| XX       | Key  | Location/Qualifiers     |
| FT       | CDS  | 7..477                  |
| FT       | CDS  | /*tag= a                |
| FT       | CDS  | /product= "prg-1/IEX-1" |
| XX       | WO200216416-A2.  |                         |
| XX       | WO200216416-A2.  |                         |
| XX       | 28-FEB-2002.   |                         |
| XX       | 28-FEB-2002.   |                         |
| XX       | 21-AUG-2001; 2001WO-US026089.  |                         |
| XX       | 21-AUG-2001; 2001WO-US026089.  |                         |
| XX       | 22-AUG-2000; 2000US-022159P.   |                         |
| XX       | 22-AUG-2000; 2000US-022159P.   |                         |
| XX       | (BGM ) BRIGHAM & WOMENS HOSPITAL INC.                                    |                         |
| XX       | (PFIZ ) PFIZER INC.  |                         |
| XX       | Lee RT, Landschulz KT, Kennedy SP, Thompson JF, Turi TG;                 |                         |
| XX       | WPI; 2002-280912/32.   |                         |
| XX       | P-PSDB; AAU78232.  |                         |
| XX       | Novel nucleic acid molecule encoding Mechanically Induced Vascular       |                         |
| XX       | Receptor-1 polypeptide, useful for treating cardiovascular diseases.     |                         |
| XX       | Example; Page 90; 105pp; English.  |                         |
| XX       | The invention relates to an isolated nucleic acid molecule encoding a    |                         |
| XX       | Mechanically induced Vascular Receptor (MIVR)-1 polypeptide having       |                         |

CC cardiac cell anti-apoptotic activity and fragments of it provided they  
 CC are not identical to Genbank sequences AI761441.1, AI594390, NM 004338  
 CC and A0777461. Also included are expression vectors, host cells, the MIVR-  
 CC 1 polypeptide, MIVR-1 binding peptides, modulators of MIVR-1, contacting  
 CC a molecule having cardiac cell anti-apoptotic activity with a candidate  
 CC agent, where the molecule is a nucleic acid molecule comprising MIVR-1,  
 CC IEX-1, VDUPL-1, BTG-2 and IIS-11d or its expression product, determining  
 CC if the anti-apoptotic activity is modulated and thereby identifying a  
 CC modulator. The cardiac cell anti-apoptotic molecules and nucleic acids  
 CC of the invention are useful for treating, diagnosing and monitoring  
 CC progression of such diseases and disorders as characterised by increased  
 CC apoptotic cell-death of vascular endothelial cells e.g. cardiac  
 CC hypertrophy, myocardial infarction, stroke, arteriosclerosis and heart  
 CC failure. The present sequence encodes human proliferation regulated gene-  
 CC 1, prg-1 (also known as IEX-1) protein which is also mechanically induced  
 CC and has an apoptosis regulatory function  
 XX  
 SQ Sequence 477 BP; 72 A; 205 C; 117 G; 83 T; 0 U; 0 Other;

Query Match 100.0%; Score 258; DB 6; Length 477;  
 Best Local Similarity 100.0%; Pred. No. 9.3e-60;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 Db 217 GTCCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 276  
 QY 61 CTCACCATGCTTCTTCTGCGAGATCCTGATGGCTGAGAGGGTGTGCCGGCCCTGCT 120  
 Db 277 CTCACCATGCTTCTTCTGCGAGATCCTGATGGCTGAGAGGGTGTGCCGGCCCTGCT 336  
 QY 121 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 180  
 Db 337 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 396  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTTCAGACATTTCTCCAG 240  
 Db 397 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTTCAGACATTTCTCCAG 456  
 QY 241 CAACACCGCGCGCCCTTC 258  
 Db 457 CAACACCGCGCGCCCTTC 474

RESULT 2  
 AAT94471  
 ID AAT94471 standard; cDNA; 1228 BP.  
 XX AC AAT94471;  
 XX  
 DT 03-MAR-1998 (first entry)  
 XX  
 DE Human Fchd605 gene differentially regulated in monocytes.  
 XX  
 KW Fchd602 gene; differential expression; monocyte; human; foam cell;  
 KW cardiovascular disease; arteriosclerosis; ischaemia; reperfusion;  
 KW hypertension; restenosis; arterial inflammation; therapy; diagnosis;  
 KW drug screening; marker; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XX Key Location/Qualifiers  
 FT CDS 1..471  
 FT /\*tag= a  
 XX  
 XX WC973C065-A1.  
 XX  
 XX 21-AUG-1997.  
 XX  
 XX 14-FEB-1997; 97WO-0002291.  
 XX  
 XX 16-FEB-1996; 96US-0011787P.  
 XX  
 XX 13-FEB-1997; 97US-00799910.

XX PA (MILL-) MILLENNIUM PHARM INC.  
 XX PI Palb DA;  
 XX WPI; 1997-424966/39.  
 XX P-PSDB; AAW36006.  
 PT New genes differentially expressed in cardiovascular disease - used for  
 PT diagnosis, drug screening and treatment of cardiovascular disease, e.g.  
 PT arteriosclerosis, restenosis, hypertension, etc.  
 XX Claim 1; Fig 5; 163pp; English.  
 CC Fchd605 is a novel human gene that is up-regulated in monocytes treated  
 CC with oxidised low density lipoproteins (LDL). Differential display was  
 CC used to detect genes that are differentially expressed in monocytes  
 CC treated so as to simulate the conditions under which foam cells develop  
 CC during atherogenesis. Both fchd605 and fchd602 (see AAT94470) are up-  
 CC regulated under the disease condition of treatment with oxidised LDL. The  
 CC fchd605 gene product (see AAW36006) has sequence similarity to mouse  
 CC giv96. The discovery of the up-regulation of these 2 genes provides a  
 CC fingerprint profile, e.g. markers for the study of cardiovascular  
 CC diseases, including arteriosclerosis, ischaemia/reperfusion, hypertension,  
 CC restenosis and arterial inflammation. Methods are provided for the  
 CC diagnosis, monitoring in clinical trials, screening for therapeutically  
 CC effective compounds, and treatment of cardiovascular diseases based on  
 CC discoveries regarding the expression patterns of novel genes fchd531 (see  
 CC AAT94467), fchd540 (see AAT94468), fchd545 (see AAT94469), fchd602 and  
 CC fchd605  
 XX

SQ Sequence 1228 BP; 265 A; 362 C; 340 G; 261 T; 0 U; 0 Other;  
 Query Match 100.0%; Score 258; DB 2; Length 1228;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-59;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 Db 211 GTCCGGCGCCAGTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 270  
 QY 61 CTCACCATGCTTCTTCTGCGAGATCCTGATGGCTGAGAGGGTGTGCCGGCCCTGCT 120  
 Db 271 CTCACCATGCTTCTTCTGCGAGATCCTGATGGCTGAGAGGGTGTGCCGGCCCTGCT 330  
 QY 121 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 180  
 Db 331 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTCCCGCTCTCGAG 390  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTTCAGACATTTCTCCAG 240  
 Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTTCAGACATTTCTCCAG 450  
 QY 241 CAACACCGCGCGCCCTTC 258  
 Db 451 CAACACCGCGCGCCCTTC 468

RESULT 3  
 AAZ50711  
 ID AAZ50711 standard; DNA; 1228 BP.  
 XX AC AAZ50711;  
 XX  
 DT 31-MAY-2000 (first entry)  
 XX  
 DE Nucleotide sequence of human fchd605 gene.  
 XX  
 KW fchd605 gene; human; cardiovascular disease; oncogenic disorder;  
 KW diabetic retinopathy; fibroproliferative disorder; arteriosclerosis;  
 KW TGF-beta signalling pathway; TGF; Transforming growth factor;  
 KW pancreatic cancer; angiogenesis; inflammation; fibrosis; tumour growth;  
 KW vascularisation; cytostatic; antidiabetic; ophthalmological; ds.

XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
XX CDS l. .471  
FT /\*tag= a  
FT /product= "fcdh605 protein"  
XX W0200006206-A1.  
XX PN 10-FEB-2000.  
XX PD 30-JUL-1999; 99WO-US017394.  
XX PR 30-JUL-1998; 98US-00126640.  
XX PA (MILL-) MILLENNIUM PHARM INC.  
XX Falb DA;  
XX WPI; 2000-205414/18.  
XX P-PSDB; AAY45017.  
XX Identifying substances for ameliorating symptoms of fibroproliferative diseases or oncogenic related disorders.  
XX Example; Fig 5; 214pp; English.  
XX The patent discloses methods for the treatment and diagnosis of cardiovascular diseases by novel human genes which are differentially expressed in different cardiovascular disease states. Compositions which can modify TGF-beta signalling pathway are identified by screening. These are used therapeutically to treat fibroproliferative and oncogenic disorders; especially TGF (Transforming growth factor)-beta related disorders, including diabetic retinopathy, arteriosclerosis, pancreatic cancer, angiogenesis, inflammation, fibrosis, tumour growth and vascularisation. The present sequence is fcdh605 gene which is up - regulated in monocytes treated with oxidised LDL (low density lipoprotein) can be used to design cardiovascular disease treatment strategies. Depending on whether the up-regulation has a pathogenic or protective effect treatment methods can be designed to increase or decrease the activity of the protein product of the gene  
XX  
XX Sequence 1228 BP; 265 A; 361 C; 341 G; 261 T; 0 U; 0 Other;  
Query Match 100.0%; Score 258; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 1.1e-59;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 211 GTCGGCGCCAGCTGCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 270  
QY 61 CTCACCATGCTTTCTGCGAGATCCGATGCGTGAAGGGTGTGCGCGCCCTGCT 120  
Db 271 CTCACCATGCTTTCTGCGAGATCCGATGCGTGAAGGGTGTGCGCGCCCTGCT 330  
QY 121 CCAGAGAGCGCCCTTAACGGCGATCCCTGGCGCCACCCCTGTGTCCTCTCGAG 180  
Db 331 CCAGAGAGCGCCCTTAACGGCGATCCCTGGCGCCACCCCTGTGTCCTCTCGAG 390  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACTTTCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTTCAGACTTTCTCCAG 450  
QY 241 CAACACCGCGCGCTTC 258  
Db 451 CAACACCGCGCGCTTC 468  
RESULT 4  
AAH34278 standard; cDNA; 1297 BP.  
ID AAH34278

XX AC AAH34278;  
XX 03-SEP-2001 (first entry)  
XX Human colon cancer antigen encoding cDNA SEQ ID NO:1360.  
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
XX colorectal carcinoma; chromosome 6; ss.  
XX Homo sapiens.  
XX W0200122920-A2.  
XX 05-APR-2001.  
XX 28-SEP-2000; 2000WO-US026524.  
XX 29-SEP-1999; 99US-0157137P.  
XX 03-NOV-1999; 99US-0163280P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX WPI; 2001-236357/24.  
XX P-PSDB; AAG74873.  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers.  
XX Claim 1; Page 3075; 9803pp; English.  
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patient's own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAG77789 represent sequences used in the exemplification of the present invention. N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922  
XX Sequence 1297 BP; 281 A; 392 C; 357 G; 265 T; 0 U; 2 Other;  
Query Match 100.0%; Score 258; DB 4; Length 1297;  
Best Local Similarity 100.0%; Pred. No. 1.2e-59;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCCAGCTGCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 242 GTCGGCGCCAGCTGCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 301  
QY 61 CTCACCATGCTTTCTGCGAGATCCGATGCGTGAAGGGTGTGCGCGCCCTGCT 120  
Db 302 CTCACCATGCTTTCTGCGAGATCCGATGCGTGAAGGGTGTGCGCGCCCTGCT 361  
QY 121 CCAGAGAGCGCCCTTAACGGCGATCCCTGGCGCCACCCCTGTGTCCTCTCGAG 180  
Db 362 CCAGAGAGCGCCCTTAACGGCGATCCCTGGCGCCACCCCTGTGTCCTCTCGAG 421  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTTCAGACTTTCTCGAG 240  
Db 422 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACTTCAGACTTTCTCGAG 481

disorders, cancer, tumours, foetal and developmental abnormalities.  
 CC haematopoietic disorders, diseases of the immune system, AIDS, autoimmune  
 CC diseases (e.g., rheumatoid arthritis), inflammation, allergies,  
 CC neurological disorders (e.g., Alzheimer's disease, Parkinson's disease),  
 CC cognitive disorders, schizophrenia, asthma, skin disorders (e.g.,  
 CC psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders,  
 CC angiogenic disorders, kidney disorders, gastrointestinal disorders,  
 CC pregnancy-related disorders, endocrine disorders, and infections. The  
 CC proteins can also be used to aid wound healing and epithelial cell  
 CC proliferation, to prevent skin aging due to sunburn, to maintain organs  
 CC before transplantation, for supporting cell culture of primary tissues,  
 CC to regenerate tissues, to identify their cognate ligands or binding  
 CC partners, and in chemotaxis, and can be used as a food additive or  
 CC preservative to modify storage properties. Antibodies specific for a  
 CC protein of the invention can be used in alleviating symptoms associated  
 CC with the disorders mentioned above, and in diagnostic immunoassays e.g.,  
 CC radioimmunoassay or enzyme linked immunosorbent assay (ELISA). The  
 CC present sequence represents a human secreted protein-encoding cDNA of the  
 CC invention  
 XX  
 SQ Sequence 1308 BP; 282 A; 398 C; 360 G; 266 T; 0 U; 2 Other;  
 Query Match 100.0%; Score 258; DB 4; Length 1308;  
 Best Local Similarity 100.0%; Pred.No.1.2e-59;  
 Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GTCCGCGCCAGCTGCAGTTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 60  
 DB 253 GTCCGCGCCAGCTGCAGTTCGAGGACCGAACCCAGCCAAAGGCTTCTTTCTGCTG 312  
 QY 61 CTCACCATGCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGCGCCCTGCCT 120  
 DB 313 CTCACCATGCTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCGGCGCCCTGCCT 372  
 QY 121 CCAGAGGAGCCCTTAACGCGCATCTCTGGCCCAACCCCTGTGTCCTCCCTCTCCAG 180  
 DB 373 CCAGAGGAGCCCTTAACGCGCATCTCTGGCCCAACCCCTGTGTCCTCCCTCTCCAG 432  
 QY 181 CCTTTTATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTCAGCACTTCTCTCCAG 240  
 DB 433 CCTTTTATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTCAGCACTTCTCTCCAG 492  
 QY 241 CAACACCCGCGCGCTTC 258  
 DB 493 CAACACCCGCGCGCTTC 510  
 RESULT 6  
 AAK73551  
 ID AAK73551 standard; DNA; 1578 BP.  
 XX AAK73551;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:28363.  
 XX  
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 17-JAN-2001; 20C1WO-US001354.  
 XX  
 PR 31-JAN-2000; 20C0US-0179065P.  
 PR 04-FEB-2000; 20C0US-0180628P.  
 PR 24-FEB-2000; 20C0US-0184664P.  
 PR 02-MAR-2000; 20C0US-0186350P.  
 PR 16-MAR-2000; 20C0US-0189874P.  
 CC

241 CAACACCCGCGCGCTTC 258  
 482 CAACACCCGCGCGCTTC 499  
 RESULT 5  
 AAD08380  
 ID AAD08380 standard; cDNA; 1308 BP.  
 XX AAD08380;  
 XX  
 DT 09-AUG-2001 (first entry)  
 XX  
 DE Human secreted protein-encoding gene 36 cDNA clone HAUC84; SEQ ID NO:46.  
 XX  
 KW Human; secreted protein; proliferative disorder; cancer; tumour;  
 KW foetal abnormality; developmental abnormality; haematopoietic disorder;  
 KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;  
 KW inflammation; allergy; neurological disorder; Alzheimer's disease;  
 KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;  
 KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;  
 KW cardiovascular disorder; angiogenic disorder; kidney disorder;  
 KW gastrointestinal disorder; pregnancy-related disorder;  
 KW endocrine disorder; infection; wound healing; vulvetry; cell culture;  
 KW chemotaxis; food additive; gene therapy; binding partner identification;  
 KW ss.  
 XX  
 OS Homo sapiens.  
 XX  
 XH Key Location/Qualifiers  
 CDS 265..513  
 FT \*tag= a  
 FT /product= "Human secreted protein precursor"  
 FT /note= "CDS does not include start codon"  
 FT /partial  
 FT sig\_peptide 265..363  
 FT \*tag= b  
 FT mat\_peptide 364..510  
 FT \*tag= c  
 FT /product= "Mature human secreted protein"  
 XX  
 EN WO200077022-A1.  
 XX  
 PD 21-DEC-2000.  
 XX  
 PF 01-JUN-2000; 2000WO-US015136.  
 XX  
 PF 11-JUN-1999; 99US-0138629P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM, Komatsoulis GA;  
 XX  
 DR WPI; 2001-367020/38.  
 DR P-PSDB; AAS03633.  
 XX  
 PT Nucleic acids encoding 50 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating diseases, e.g. Parkinson's  
 PT disease, botulism, cancers and Scimitar syndrome.  
 XX  
 XX Claim 1; Page 515-516; 614pp; English.  
 XX  
 CC AAD08345-AAD08394 represent cDNAs corresponding to 50 human secreted  
 CC protein genes and AAE03898-AAE03947 represent the proteins they encode.  
 CC AAE03948-AAE03996 represent human secreted protein fragments or variants.  
 CC The genes and their secreted proteins are useful for preventing, treating  
 CC or ameliorating medical conditions, e.g., by protein or gene therapy.  
 CC Pathological conditions can be diagnosed by determining the amount of the  
 CC new protein in a sample or by determining the presence of mutations in  
 CC the new genes. Specific uses are described for each of the 50 genes,  
 CC based on the tissues in which they are most highly expressed, and include  
 CC developing products for the diagnosis or treatment of proliferative

PR 17-MAR-2000; 2000US-0190076P.  
 PR 18-APR-2000; 2000US-0198123P.  
 PR 19-MAY-2000; 2000US-0205515P.  
 PR 19-JUN-2000; 2000US-0209467P.  
 PR 28-JUN-2000; 2000US-0214886P.  
 PR 30-JUN-2000; 2000US-0215135P.  
 PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 14-JUL-2000; 2000US-0217496P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225757P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226275P.  
 PR 22-AUG-2000; 2000US-0226661P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 21-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 25-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 29-SEP-2000; 2000US-0235836P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239935P.

PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241211P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX  
 XX Nucleic acids encoding human immune/haematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
 PT  
 XX Disclosure; SEQ ID NO 28363; 3071pp + Sequence Listing; English.  
 PS  
 XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)

CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen generic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 1578 BP; 303 A; 490 C; 445 G; 340 T; 0 U; 0 Other;

Query Match 100.0%; Score 258; DB 4; Length 1578;  
Best Local Similarity 100.0%; Pred. No. 1.2e-59;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCGAGCTGCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTCTTCGCTG 60  
Db 345 GTCGGCGCGAGCTGCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTCTTCGCTG 404  
QY 61 CTACCATCGTCTTCTGCGAGATCTGATGCTGAGAGGCTGTCGGGCGCCCTGCT 120  
Db 405 CTACCATCGTCTTCTGCGAGATCTGATGCTGAGAGGCTGTCGGGCGCCCTGCT 464  
QY 121 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTGCCCGCTCTCGAG 180  
Db 465 CCAGAGAGCGCCCTTAAGCGCGCATCCCTGGCGCCACCCCTGTGTGCCCGCTCTCGAG 524  
QY 181 CCCTTAACTGACTTGGAGCCCTCGGACTGAGCTCTGAGCTCAGCACTTTCCTCGAG 240  
Db 525 CCCTTAACTGACTTGGAGCCCTCGGACTGAGCTCTGAGCTCAGCACTTTCCTCGAG 584  
QY 241 CAACACCGCGCGCTTC 258  
Db 585 CAACACCGCGCGCTTC 602

RESULT 7  
AAK73553  
ID AAK73553 standard; DNA; 1578 BP.  
XX AC AAK73553;  
XX DT 06-NOV-2001 (first entry)  
XX DE Human: immune/haematopoietic antigen genomic sequence SEQ ID NO:28365.  
XX KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX KW cytostatic; gene therapy; vaccine; metastasis; ds.  
XX OS Homo sapiens.  
XX PN WO200157182-A2.  
XX PD 09-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US001354.  
XX PR 31-JAN-2000; 2000US-0179065P.  
XX PR 04-FEB-2000; 2000US-0180628P.  
XX PR 24-FEB-2000; 2000US-0184664P.  
XX PR 02-MAR-2000; 2000US-0186350P.  
XX PR 16-MAR-2000; 2000US-0189874P.  
XX PR 17-MAR-2000; 2000US-0190076P.  
XX PR 18-APR-2000; 2000US-0196123P.  
XX PR 19-MAY-2000; 2000US-0205515P.  
XX PR 07-JUN-2000; 2000US-0209467P.  
XX PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 14-AUG-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.  
PR 06-SEP-2000; 2000US-0230438P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241786P.

PR 20-OCT-2000; 2000US-02411787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246474P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246609P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.  
PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 05-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Disclosure; SEQ ID NO 28365; 3071pp + Sequence Listing; English.  
XX  
XX AAK64951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to

CC supplement the patient's own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting the  
CC nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent  
CC disease and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK6494 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
CC represent sequences used in the exemplification of the present invention  
XX  
XX SQ Sequence 1578 BP; 303 A; 490 C; 445 G; 340 T; 0 U; 0 Other;  
Query Match 100.0%; Score 258; DB 4; Length 1578;  
Best Local Similarity 100.0%; Pred. No. 1.2e-59;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCCGGCGCCAGTCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTTCTGCTG 60  
DB 345 GTCCGGCGCCAGTCCAGTCGAGGAAACCGAACCCAGCCAAAGGCTTCTTCTGCTG 404  
QY 61 CTCACCATCGTCTTCTGCGCAGATCCTGATGGCTGAGAGAGGGTGTGCGCGCGCCCTGCT 120  
DB 405 CTCACCATCGTCTTCTGCGCAGATCCTGATGGCTGAGAGAGGGTGTGCGCGCGCCCTGCT 464  
QY 121 CCAGAGAGCGCCCTTAACGCGCGATCCCTCGCGCCACCCCTGTCCTGCTGCTGAG 180  
DB 465 CCAGAGAGCGCCCTTAACGCGCGATCCCTCGCGCCACCCCTGTCCTGCTGCTGAG 524  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCAG 240  
DB 525 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGGACCTCAGCACTTCTCCAG 584  
QY 241 CAACACCGCGCGCCCTTC 258  
DB 585 CAACACCGCGCGCCCTTC 602  
RESULT 8  
AAL37434  
ID AAL37434 standard; DNA; 1578 BP.  
XX  
XX AC AAL37434;  
XX  
XX DT 08-JAN-2002 (first entry)  
XX  
XX DE Human musculoskeletal system related polynucleotide SEQ ID NO 3799.  
XX  
XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
XX atiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;  
XX vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
XX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
XX neurological disease; infection; human; secreted protein;  
XX musculoskeletal system; ds.  
XX  
XX OS Homo sapiens.  
XX  
XX PN WO200155367-A1.  
XX  
XX PD 02-AUG-2001.  
XX  
XX PF 17-JAN-2001; 2001WO-US001338.  
XX  
XX PR 31-JAN-2000; 2000US-0179065P.  
XX PR 04-FEB-2000; 2000US-0180628P.  
XX PR 24-FEB-2000; 2000US-0184664P.  
XX PR 02-MAR-2000; 2000US-0186350P.  
XX PR 16-MAR-2000; 2000US-0189874P.  
XX PR 17-MAR-2000; 2000US-0190076P.  
XX PR 18-APR-2000; 2000US-0198123P.  
XX PR 19-MAY-2000; 2000US-0205515P.  
XX PR 07-JUN-2000; 2000US-0209467P.  
XX PR 28-JUN-2000; 2000US-0214886P.  
XX PR 30-JUN-2000; 2000US-0215135P.



PR 07-JUL-2000; 2000US-0216647P.  
 PR 07-JUL-2000; 2000US-0216880P.  
 PR 11-JUL-2000; 2000US-0217487P.  
 PR 11-JUL-2000; 2000US-0217496P.  
 PR 14-JUL-2000; 2000US-0218290P.  
 PR 26-JUL-2000; 2000US-0220963P.  
 PR 26-JUL-2000; 2000US-0220964P.  
 PR 14-AUG-2000; 2000US-0224518P.  
 PR 14-AUG-2000; 2000US-0224519P.  
 PR 14-AUG-2000; 2000US-0225213P.  
 PR 14-AUG-2000; 2000US-0225214P.  
 PR 14-AUG-2000; 2000US-0225266P.  
 PR 14-AUG-2000; 2000US-0225267P.  
 PR 14-AUG-2000; 2000US-0225268P.  
 PR 14-AUG-2000; 2000US-0225270P.  
 PR 14-AUG-2000; 2000US-0225447P.  
 PR 14-AUG-2000; 2000US-0225575P.  
 PR 14-AUG-2000; 2000US-0225758P.  
 PR 14-AUG-2000; 2000US-0225759P.  
 PR 18-AUG-2000; 2000US-0226279P.  
 PR 22-AUG-2000; 2000US-0226681P.  
 PR 22-AUG-2000; 2000US-0226868P.  
 PR 22-AUG-2000; 2000US-0227182P.  
 PR 23-AUG-2000; 2000US-0227009P.  
 PR 30-AUG-2000; 2000US-0228924P.  
 PR 01-SEP-2000; 2000US-0229287P.  
 PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229509P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231444P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 08-SEP-2000; 2000US-0232081P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 14-SEP-2000; 2000US-0232397P.  
 PR 14-SEP-2000; 2000US-0232398P.  
 PR 14-SEP-2000; 2000US-0232399P.  
 PR 14-SEP-2000; 2000US-0232400P.  
 PR 14-SEP-2000; 2000US-0232401P.  
 PR 14-SEP-2000; 2000US-0232633P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235835P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239933P.  
 PR 13-OCT-2000; 2000US-0239937P.  
 PR 20-OCT-2000; 2000US-0240560P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241787P.

PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241826P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246609P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-451937/48.  
 XX  
 XX Isolated polypeptide for treating, preventing and/or prognosing  
 PT disorders related to the musculoskeletal system including musculoskeletal  
 PT cancers and also for testing and detection e.g. diagnosis.  
 XX  
 PS Example 2; SEQ ID NO 3799; 781pp + Sequence Listing; English.  
 XX  
 CC The invention relates to novel genes (AAL34669-AAL37666) and proteins  
 CC (AB03087-AB04109) associated with the musculoskeletal system useful for  
 CC preventing, treating or ameliorating medical conditions e.g. by protein  
 CC or gene therapy. The genes are isolated from a range of human tissues  
 CC disclosed in the specification. The nucleic acids, proteins, antibodies  
 CC and (ant)agonists are useful in the diagnosis, treatment and prevention  
 CC of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the  
 CC adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver,





XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0218290P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 22-AUG-2000; 2000US-0225758P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 05-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239935P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241785P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
XX (ROSE/) ROSEN C A.  
XX (RUBE/) RUBEN S M.  
XX (BARA/) BARASH S C.  
XX Rosen CA, Ruben SM, Barash SC;  
XX WPI; 2003-128199/12.  
XX  
XX Isolated nucleic acid molecules encoding musculoskeletal system  
XX associated polypeptides, useful for detecting disorders, e.g. cancer.  
XX  
XX Disclosure; SEQ ID NO 3799; 321pp; English.  
XX  
XX The invention describes an isolated nucleic acid molecule comprising a  
XX sequence encoding musculoskeletal system associated polypeptides useful  
XX for detecting disorders, e.g., cancer or cancer metastases, in animals or  
XX humans. The nucleic acid: stimulates re-vascularisation of ischemic  
XX tissues associated with conditions such as thrombosis, arteriosclerosis,  
XX and other cardiovascular conditions; treats wounds due to injuries,  
XX burns, post-operative tissue repair, and ulcers; stimulates angiogenesis  
XX and limb regeneration; stimulates neuronal growth; can treat and prevent

CC neuronal damage occurring in certain disorders or neurodegenerative  
CC conditions, such as, Alzheimer's disease, Parkinson's disease, and AIDS-  
CC related complex; stimulates chondrocyte growth, thus they can be used to  
CC enhance bone and periodontal regeneration and aid in tissue transports or  
CC bone grafts; prevents skin aging due to sunburn by stimulating  
CC keratinocyte growth; prevents hair loss, since FGF family members  
CC activate hair-forming cells and promotes melanocyte growth; stimulates  
CC growth and differentiation of hematopoietic cells and bone marrow cells  
CC when used in combination with other cytokines; maintains organs before  
CC transplantation or for supporting cell culture of primary tissues;  
CC induces tissue of mesodermal origin to differentiate in early embryos;  
CC increases or decreases the differentiation or proliferation of embryonic  
CC stem cells, besides, haematopoietic lineage; modulates mammalian  
CC characteristics, such as, body height, weight, hair colour, eye colour,  
CC skin, percentage of adipose tissue, pigmentation, size, and shape (e.g.,  
CC cosmetic surgery); modulates mammalian metabolism; changes mammal's metal  
CC state or physical state by influencing biorhythms, cardiac rhythms,  
CC depression, tendency for violence, tolerance for pain, reproductive,  
CC capabilities, hormonal or endocrine levels, appetite, libido, memory, or  
CC stress; increases or decreases storage capabilities, fat content, lipid,  
CC protein, carbohydrate, vitamins, minerals, cofactors or other nutritional  
CC components. This sequence encodes a novel human musculoskeletal system  
CC antigen. Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from the US patent office at  
CC ftp.segdata.uspto.gov/sequence.html?DocID=20020147140  
XX

SQ Sequence 1578 BP; 303 A; 490 C; 445 G; 340 T; 0 U; 0 Other;

Query Match 100.0%; Score 258; DB 7; Length 1578;

Best Local Similarity 100.0%; Pred. No. 1.2e-59;

Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTCTCTG 60

Db 345 GTCCGGCGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTCTCTG 404

QY 61 CTACCATGTCTTCTGCGAGATCTGATGGCTGAGAGGGTGTGCGCGCCCTGCCT 120

Db 405 CTACCATGTCTTCTGCGAGATCTGATGGCTGAGAGGGTGTGCGCGCCCTGCCT 464

QY 121 CCAGAGAGCGCCCTAACGGCGCATCCCTGGCGCCACCCCTGTGCCCCGCTCTCGAG 180

Db 465 CCAGAGAGCGCCCTAACGGCGCATCCCTGGCGCCACCCCTGTGCCCCGCTCTCGAG 524

QY 181 CCCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGAACCTCAGCACTTTCTCCAG 240

Db 525 CCCTTTAATCTGACTTGGAGCCCTCGGACTACGCTCTGGAACCTCAGCACTTTCTCCAG 584

QY 241 CACACCCGCGCGCTTC 258

Db 585 CACACCCGCGCGCTTC 602

RESULT 11

ABX60424

ID ABX60424 standard; cDNA; 1578 BP.

XX AC ABX60424;

XX 26-FEB-2003 (first entry)

XX cDNA encoding novel human musculoskeletal system antigen #2768.

XX Gene; ss; musculoskeletal system antigen; cancer; metastasis;  
XX re-vascularisation; thrombosis; arteriosclerosis; mineral content;  
XX cardiovascular condition; wound; injury; burn; angiogenesis; ulcer;  
XX post-operative tissue repair; limb regeneration; neuronal growth;  
XX neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
XX AIDS-related complex; chondrocyte growth; bone regeneration;  
XX periodontal regeneration; tissue transport; bone graft; skin aging;  
XX keratinocyte growth; hair loss; melanocyte growth; cell proliferation;  
XX cell growth; organ transplant; cell differentiation; body height; weight;





CC of BPH cells or BPH-like cell population, exposing the cells to the  
 CC agent, preparing a second gene expression profile of the agent exposed  
 CC cells, and comparing the first and second gene expression profiles. (I)  
 CC is useful for diagnosing the onset or progression of BPH. (II) is useful  
 CC for identifying an agent that modulates the onset or progression of BPH.  
 CC The methods are useful to present information identifying the expression  
 CC level in a tissue or cells, by comparing the expression level of genes  
 CC given in the specification in the tissue or cells to the level of  
 CC expression of gene in the database, and displaying the expression levels  
 CC of at least one gene in the tissue or cell sample compared to the  
 CC expression level in BPH. Agents using (II) are useful for treating BPH or  
 CC prostate cancer. ABK4106-ABK54860 represent human benign prostatic  
 CC hyperplasia gene sequences of the invention

XX  
 SQ Sequence 1223 BP; 249 A; 368 C; 342 G; 264 T; 0 U; 0 Other;  
 Query Match 99.4%; Score 256.4; DB 6; Length 1223;  
 Best Local Similarity 99.6%; Pred. No. 3.1e-59;  
 Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTCGGCGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTCTG 60  
 DB 229 GTCGGCGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTCTG 288  
 QY 61 CTCACCATCTCTTCCGAGATCCTGATGCTGAGAGGGTGTCCGCGCCCTGCT 120  
 DB 289 CTCACCATCTCTTCCGAGATCCTGATGCTGAGAGGGTGTCCGCGCCCTGCT 348  
 QY 121 CCAGAGAGCGCCCTTAACGCGCATCCCTGGCGGCCACCCCTGTGTCCCGCTCTCGAG 180  
 DB 349 CCAGAGAGCGCCCTTAACGCGCATCCCTGGCGGCCACCCCTGTGTCCCGCTCTCGAG 408  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGAGCACTTTCCTCCAG 240  
 DB 409 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGAGCACTTTCCTCCAG 468  
 QY 241 CAACACCGCGCGCTTC 258  
 DB 469 CAACACCGCGCGCTTC 486

RESULT 14  
 ABN96540  
 ID ABN96540 standard; DNA; 1223 BP.  
 XX  
 AC ABN96540;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Gene #3038 used to diagnose liver cancer.  
 XX  
 KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;  
 KW metastatic liver tumour; cytostatic; expression profile; disease state;  
 KW disease progression; drug toxicity; drug efficacy; drug metabolism.  
 XX  
 OS Homo sapiens.  
 XX  
 XN WO200229103-A2.  
 XX  
 PD 11-APR-2002.  
 XX  
 PF 02-OCT-2001; 2001WO-US030589.  
 XX  
 PR 02-OCT-2000; 2000US-0237054P.  
 XX  
 XX (GENE-) GENE LOGIC INC.  
 XX  
 XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;  
 XX WPI; 2002-426119/45.  
 XX  
 XX Diagnosing and detecting the progression of liver cancer, hepatocellular  
 PT carcinoma or metastatic liver tumor in a patient, involves detecting the

PT level of expression of two or more genes in a liver tissue sample.  
 XX  
 PS Claim 1; SEQ ID NO 3038; 298pp; English.  
 XX  
 CC The invention relates to a novel method for diagnosing and detecting the  
 CC progression of liver cancer, hepatocellular carcinoma or metastatic liver  
 CC tumour in a patient, and differentiating metastatic liver cancer from  
 CC hepatocellular carcinoma in a patient, involving detecting the level of  
 CC expression of two or more genes represented in ABN93503-ABN97455 in a  
 CC tissue sample. The method of the invention has hepatotropic, and  
 CC cytostatic activity. The method is useful for diagnosing and detecting  
 CC the progression of liver cancer, hepatocellular carcinoma and metastatic  
 CC liver carcinoma in a patient. The method is useful for identifying  
 CC expression profiles which serve as useful diagnostic markers as well as  
 CC markers that can be used to monitor disease states, disease progression,  
 CC drug toxicity, drug efficacy and drug metabolism. Note: The sequence data  
 CC for this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pot\_sequences

XX  
 SQ Sequence 1223 BP; 249 A; 368 C; 342 G; 264 T; 0 U; 0 Other;  
 Query Match 99.4%; Score 256.4; DB 6; Length 1223;  
 Best Local Similarity 99.6%; Pred. No. 3.1e-59;  
 Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCGGCGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTCTG 60  
 DB 229 GTCGGCGCCAGCTCCAGTCGAGAACCGAACCCAGCCAAAGGCTTCTTTCTCTG 288  
 QY 61 CTCACCATCTCTTCCGAGATCCTGATGCTGAGAGGGTGTCCGCGCCCTGCT 120  
 DB 289 CTCACCATCTCTTCCGAGATCCTGATGCTGAGAGGGTGTCCGCGCCCTGCT 348  
 QY 121 CCAGAGAGCGCCCTTAACGCGCATCCCTGGCGGCCACCCCTGTGTCCCGCTCTCGAG 180  
 DB 349 CCAGAGAGCGCCCTTAACGCGCATCCCTGGCGGCCACCCCTGTGTCCCGCTCTCGAG 408  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGAGCACTTTCCTCCAG 240  
 DB 409 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTCGAGCACTTTCCTCCAG 468  
 QY 241 CAACACCGCGCGCTTC 258  
 DB 469 CAACACCGCGCGCTTC 486

RESULT 15  
 AAH23085  
 ID AAH23085 standard; DNA; 1232 BP.  
 XX  
 AC AAH23085;  
 XX  
 DT 17-SEP-2001 (first entry)  
 XX  
 DE Osteoarthritis tissue-derived nucleic acid sequence #15.  
 XX  
 KW Osteoarthritis; infectious disorder; Crohn's disease; sepsis; human;  
 KW wound healing; osteopathic; anti-arthritis; anti-inflammatory; vulnerary;  
 KW antibacterial; antiallergic; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 XN WO200153531-A2.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 18-JAN-2001; 2001WO-US000016.  
 XX  
 PR 18-JAN-2000; 2000US-0176523P.  
 XX  
 XX (PFAA ) PHARMACIA CORP.  
 XX

PI Puippard D, Vasantbakamur G, Dorson S, Ma X;  
XX WPI; 2001-451914/48.  
DR  
XX  
PT Substantially purified protein, polypeptide or their fragments, used to  
PT identify a biologically active compound or composition and treat  
PT mammalian osteoarthritis.  
XX  
XX  
PS Claim 1; Page 102; 144pp; English.  
PS  
CC Sequences AM23071-23152 represent nucleic acid sequences derived from  
CC osteoarthritis tissues. The sequences are useful as probes and for the  
CC diagnosis or prognosis of mammalian osteoarthritis. The polynucleotides  
CC and polypeptides of the invention are useful for generating diagnostic  
CC reagents, as targets for small molecule drug development, generation of  
CC therapeutics, and cloning genes. Specific antibodies are used to generate  
CC enzyme linked immunosorbent assays for detection of osteoarthritis. The  
CC invented molecules can be used to treat osteoarthritis or to analyse the  
CC disease-modifying activity of osteoarthritis drugs. Other disorders  
CC treatable using the nucleic acid sequences include atopic, inflammatory  
CC and infectious disorders e.g. Crohn's disease and sepsis, and wound  
CC healing  
XX  
SQ Sequence 1232 BP; 251 A; 371 C; 345 G; 265 T; 0 U; 0 Other;  
Query Match 91.5%; Score 236; DB 4; Length 1232;  
Best Local Similarity 99.2%; Pred.No.9.4e-54;  
Matches 258; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
Qy 1 GTCCGGCCGAGTCCAGTCGAGGAAACCGAACCCAGACCAAAAGGCTTCTTTTCGCTG 60  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
232 GTCCGGCCGAGTCCAGTCGAGGAAACCGAACCCAGACCAAAAGGCTTCTTTTCGCTG 291  
Qy 61 CTCACCATGCTCTTCTGCGCATCTCTGATGGCTGAAGAGGCTGCGGCGCCCTCCCT 120  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
292 CTCACCATGCTCTTCTGCGCATCTCTGATGGCTGAAGAGGCTGCGGCGCCCTCCCT 351  
Qy 121 CC-AGAGAGAGCGCCCTTAACGGCGCATCCCT-GCGCGCCACCCCTGTGTCCCGCTCG 178  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
352 CCAAGAGAGAGCGCCCTTAACGGCGCATCCCTGTGGCGCCACCCCTGTGTCCCGCTCG 411  
Qy 179 AGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCACTTCTCTCC 238  
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
412 AGCCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTTGGACCTCAGCACTTCTCTCC 471  
Qy 239 AGCAACACCGCGCGCTTC 258  
Db ||||||||||||||||||||  
472 AGCAACACCGCGCGCTTC 491

Search completed: March 3, 2004, 19:12:11  
Job time : 263 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

QM nucleic - nucleic search, using sw model

Run on: March 3, 2004, 18:57:48 ; Search time 58 Seconds  
(without alignments)  
2468.575 Million cell updates/sec

Title: US-08-799-910-9\_COPY\_211\_468

Perfect score: 258

Sequence: 1 GTCGCGGCCAGTGCACCT.....AGCAACACCCGCGCGCTTC 258

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:\*

- 1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | Query Match | Length  | ID | Description                            |
|------------|-------|-------------|---------|----|--|
| 1          | 258   | 100.0       | 1228    | 3  | US-08-826-246-9 Sequence 9, Appli      |
| 2          | 258   | 100.0       | 1228    | 3  | US-08-944-495-9 Sequence 9, Appli      |
| 3          | 258   | 100.0       | 1228    | 3  | US-09-126-640-5 Sequence 5, Appli      |
| 4          | 258   | 100.0       | 1228    | 3  | US-08-925-588-9 Sequence 9, Appli      |
| 5          | 258   | 100.0       | 1228    | 4  | US-09-288-292A-5 Sequence 5, Appli     |
| 6          | 258   | 100.0       | 1228    | 4  | US-09-372-044-9 Sequence 9, Appli      |
| 7          | 258   | 100.0       | 1228    | 4  | US-08-825-486-9 Sequence 9, Appli      |
| 8          | 43    | 16.7        | 7218    | 1  | US-08-232-463-14 Sequence 14, Appli    |
| 9          | 39.6  | 15.3        | 4403765 | 3  | US-09-103-840A-2 Sequence 2, Appli     |
| 10         | 39.6  | 15.3        | 4411529 | 3  | US-09-103-840A-1 Sequence 1, Appli     |
| 11         | 39.4  | 15.3        | 2847    | 1  | US-08-087-007-2 Sequence 2, Appli      |
| 12         | 39.4  | 15.3        | 2847    | 3  | US-08-483-433-2 Sequence 2, Appli      |
| 13         | 39.4  | 15.3        | 2847    | 5  | PCT-US92-05920-2 Sequence 2, Appli     |
| 14         | 38.6  | 15.0        | 6354    | 3  | US-09-058-389A-5 Sequence 5, Appli     |
| 15         | 38.6  | 15.0        | 6354    | 4  | US-09-611-781-5 Sequence 5, Appli      |
| 16         | 37.6  | 14.6        | 170     | 3  | US-09-058-389A-13 Sequence 13, Appli   |
| 17         | 37.6  | 14.6        | 170     | 4  | US-09-611-781-13 Sequence 13, Appli    |
| 18         | 34.8  | 13.5        | 2953    | 2  | US-08-859-201-1 Sequence 1, Appli      |
| 19         | 34.8  | 13.5        | 2953    | 4  | US-09-402-214-1 Sequence 1, Appli      |
| 20         | 34.2  | 13.3        | 391     | 4  | US-09-621-976-8414 Sequence 1, Appli   |
| 21         | 33.8  | 13.1        | 423     | 4  | US-09-252-991A-10390 Sequence 10990, A |
| 22         | 33.8  | 13.1        | 786     | 4  | US-09-252-991A-10398 Sequence 10928, A |
| 23         | 33.8  | 13.1        | 1032    | 4  | US-09-252-991A-11060 Sequence 11060, A |
| 24         | 33.8  | 13.1        | 1152    | 4  | US-09-489-039A-2755 Sequence 2755, Ap  |
| 25         | 33.6  | 13.0        | 1120    | 1  | US-08-188-582-8 Sequence 8, Appli      |
| 26         | 33.6  | 13.0        | 1120    | 1  | US-08-646-715-8 Sequence 8, Appli      |
| 27         | 33.4  | 12.9        | 1392    | 4  | US-09-071-252-5 Sequence 5, Appli      |

Sequence 13635, A  
Sequence 13809, A  
Sequence 17, Appl  
Sequence 1977, Ap  
Sequence 70, Appl  
Sequence 273, Appl  
Sequence 7702, Ap  
Sequence 7861, Ap  
Sequence 236, App  
Sequence 9834, Ap  
Sequence 10150, A  
Sequence 9926, Ap  
Sequence 5924, Ap  
Sequence 7, Appli  
Sequence 3, Appli  
Sequence 12655, A  
Sequence 13312, A

28 33.4 12.9 2097 4 US-09-252-991A-13635  
29 33.4 12.9 2331 4 US-09-252-991A-13809  
30 33.4 12.9 3762 3 US-09-318-448-17  
31 33.2 12.9 1446 4 US-09-489-039A-1977  
32 32.8 12.7 513 4 US-09-216-393B-70  
33 32.8 12.7 514 4 US-09-216-393B-273  
34 32.8 12.7 1371 4 US-09-252-991A-7702  
35 32.8 12.7 1401 4 US-09-252-991A-7861  
36 32.8 12.7 1935 4 US-09-620-312D-236  
37 32.6 12.6 645 4 US-09-252-991A-9834  
38 32.6 12.6 1650 4 US-09-252-991A-10150  
39 32.6 12.6 1725 4 US-09-252-991A-9926  
40 32.4 12.6 774 4 US-09-252-991A-5924  
41 32.4 12.6 2681 3 US-08-928-213B-7  
42 32.4 12.6 43950 4 US-09-735-934A-3  
43 32.4 12.6 43950 4 US-10-060-332-3  
44 32.2 12.5 1455 4 US-09-252-991A-12655  
45 32.2 12.5 1488 4 US-09-252-991A-13312

## ALIGNMENTS

RESULT 1  
US-08-826-246-9  
; Sequence 9, Application US/08826246  
; Patent No. 6048709  
; GENERAL INFORMATION:  
; APPLICANT: Faib, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
; TITLE OF INVENTION: THE TREATMENT AND DIAGNOSIS OF  
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/826,246  
; FILING DATE: 28-MAR-1997  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: 13-FEB-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/011,787  
; FILING DATE: 16-FEB-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-078-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1228 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...468



Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGACACTTTCCTCCAG 450  
Qy 241 CAACACCCGCGCCTTC 258  
Db 451 CAACACCCGCGCCTTC 468

## RESULT 4

US-08-925-588-9  
; Sequence 9, Application US/08925588  
; Patent No. 6221628  
; GENERAL INFORMATION:  
; APPLICANT: Falb, Dean  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: PENNIE & EDMONDS LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/925,588  
; FILING DATE: 08-Sep-1997  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/799,910  
; FILING DATE: <unknown>  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7853-067-999  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)7909090  
; TELEFAX: (212)8699741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1228 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 1...468  
; OTHER INFORMATION:  
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:  
US-08-925-588-9

Query Match 100.0%; Score 258; DB 3; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 3.2e-61;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTCGGGGCCAGCTCCAGTCAGGAAACGAAACCCAGCCAAAGAGCTTCTCTTCTGCTG 60  
Db 211 GTCGGGGCCAGCTCCAGTCAGGAAACGAAACCCAGCCAAAGAGCTTCTCTTCTGCTG 270  
Qy 61 CTCACCATCGTCTTCTGCGAGATCCTGATGCTGAAGAGGCTGCGCGGCCCTGCT 120  
Db 271 CTCACCATCGTCTTCTGCGAGATCCTGATGCTGAAGAGGCTGCGCGGCCCTGCT 330  
Qy 121 CCAGAGAGCCCTTAACGCGCATCCCTGCGGCCCAACCCCTGTGTCCTGCTGAG 180  
Db 331 CCAGAGAGCCCTTAACGCGCATCCCTGCGGCCCAACCCCTGTGTCCTGCTGAG 390

Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGACACTTTCCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGACACTTTCCTCCAG 450  
Qy 241 CAACACCCGCGCCTTC 258  
Db 451 CAACACCCGCGCCTTC 468

## RESULT 5

US-09-288-292A-5  
; Sequence 5, Application US/09288292A  
; Patent No. 6359194  
; GENERAL INFORMATION:  
; APPLICANT: Dean A. Falb  
; APPLICANT: Katherine Galvin  
; APPLICANT: Michael Donovan  
; APPLICANT: Dennis Huszar  
; APPLICANT: Michael A. Gimbrone, Jr.  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and Diagnosis of Cardiovascular Disease  
; FILE REFERENCE: 7853-140-999  
; CURRENT APPLICATION NUMBER: US/09/288,292A  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: 08/870,434  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 08/799,910  
; PRIOR FILING DATE: 1997-02-13  
; PRIOR APPLICATION NUMBER: 60/011,787  
; PRIOR FILING DATE: 1996-02-16  
; PRIOR APPLICATION NUMBER: 08/485,573  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 08/386,844  
; PRIOR FILING DATE: 1995-02-10  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1228  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(468)  
US-09-288-292A-5

Query Match 100.0%; Score 258; DB 4; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 3.2e-61;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 GTCGGGGCCAGCTCCAGTCAGGAAACGAAACCCAGCCAAAGAGCTTCTCTTCTGCTG 60  
Db 211 GTCGGGGCCAGCTCCAGTCAGGAAACGAAACCCAGCCAAAGAGCTTCTCTTCTGCTG 270  
Qy 61 CTCACCATCGTCTTCTGCGAGATCCTGATGCTGAAGAGGCTGCGCGGCCCTGCT 120  
Db 271 CTCACCATCGTCTTCTGCGAGATCCTGATGCTGAAGAGGCTGCGCGGCCCTGCT 330  
Qy 121 CCAGAGAGCCCTTAACGCGCATCCCTGCGGCCCAACCCCTGTGTCCTGCTGAG 180  
Db 331 CCAGAGAGCCCTTAACGCGCATCCCTGCGGCCCAACCCCTGTGTCCTGCTGAG 390  
Qy 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGACACTTTCCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGAGACCTCAGACACTTTCCTCCAG 450  
Qy 241 CAACACCCGCGCCTTC 258  
Db 451 CAACACCCGCGCCTTC 468

## RESULT 6

US-09-372-044-9





```

1 NUMBER OF SEQUENCES: 4
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: Kilpatrick & Cody
4 STREET: 1100 Peachtree Street, Suite 2800
5 CITY: Atlanta
6 STATE: Georgia
7 COUNTRY: U.S.
8 ZIP: 30309-4530
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: PatentIn Release #1.0, Version #1.25
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/08/087,007
16 FILING DATE: 19930701
17 CLASSIFICATION: 435
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Pabst, Patrea L.
20 REGISTRATION NUMBER: 31,284
21 REFERENCE/DOCKET NUMBER: OMRF135
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: 404-815-6500
24 TELEFAX: 404-815-6555
25 INFORMATION FOR SEQ ID NO: 2:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 2847 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: single
30 TOPOLOGY: linear
31 MOLECULE TYPE: cDNA
32 HYPOTHEICAL: NO
33 ANTI-SENSE: NO
34 ORIGINAL SOURCE:
35 ORGANISM: Homo sapiens
36 IMMEDIATE SOURCE:
37 LIBRARY: GenBank HUMDAF; HUMDAFC1
38 CLONE: Human DAF cDNA
39 FEATURE:
40 NAME/KEY: misc feature
41 LOCATION: 1..819
42 OTHER INFORMATION: /note="HUMDAFC1 (Promotor and 5'
43 OTHER INFORMATION: end of Exon 1, genomic sequence)"
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

```

STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: U.S.  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,433  
FILING DATE:  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/087,007  
FILING DATE: July 1, 1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/906,394  
FILING DATE: June 29, 1992  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: OMRF135cip2 div  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404-873-8794  
TELEFAX: 404-873-9794  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2847 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULAR TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
LIBRARY: GenBank HUMDAF; HUMDAFC1  
CLONE: Human DAF CDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..819  
OTHER INFORMATION: /note= "HUMDAFC1 (Promotor a  
OTHER INFORMATION: end of Exon 1, genomic sequ  
US-08-483-433-2



/ CITY: Hackensack  
/ STATE: New Jersey  
/ COUNTRY: USA  
/ ZIP: 07601  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: PatentIn Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/09/611,781  
/ FILING DATE:  
/ CLASSIFICATION:  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: US 09/058,389  
/ FILING DATE: April 9, 1998  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Jackson Esq., David A.  
/ REGISTRATION NUMBER: 26,742  
/ REFERENCE/DOCKET NUMBER: 1340-1-013N  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 201-487-5800  
/ TELEFAX: 201-343-1684  
/ INFORMATION FOR SEQ ID NO: 5:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 6354 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: double  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: DNA (genomic)  
/ HYPOTHETICAL: NO  
/ US-09-611-781-5

Query Match 15.0%; Score 38.6; DB 4; Length 6354;  
Best Local Similarity 65.9%; Pred. No. 0.21;  
Matches 56; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
Qy 100 GGTCGTGCGCGCCCTGCTCCAGAGGAGCGCCCTAACGCGCATCCCTGGCGCCGAC 159  
Db 2461 GGTCAGAGGCGCTGCGCCCTGGCTCTCGCCCTCTGCCGAGGAGCTTCATTGAGCCCTCC 2520  
Qy 160 CCTGTGTCCCGTCTCTCGAGCCCT 184  
Db 2521 CCTGCGCCCGCTGCGCCCTCCAGCCCT 2545

Search completed: March 3, 2004, 20:07:12  
Job time : 72 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 3, 2004, 19:39:19 ; Search time 215 Seconds

(without alignments)

4388.918 Million cell updates/sec

Title: US-08-799-910-9 COPY 211 468

Perfect score: 258

Sequence: 1 GTCCGGCGCCAGCTGCCAGT.....AGCAACACCCGGCCGCTTC 258

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2421054 seqs, 1828716029 residues

Total number of hits satisfying chosen parameters: 4842108

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

```

Database : Published Applications NA.*
1: /cgn2_6/pdata/1/pubnpa/US07_PUBCOMB.seq.*
2: /cgn2_6/pdata/1/pubnpa/PCT_NEW_PUB.seq.*
3: /cgn2_6/pdata/1/pubnpa/US06_NEW_PUB.seq.*
4: /cgn2_6/pdata/1/pubnpa/US06_PUBCOMB.seq.*
5: /cgn2_6/pdata/1/pubnpa/US07_NEW_PUB.seq.*
6: /cgn2_6/pdata/1/pubnpa/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/pdata/1/pubnpa/US08_NEW_PUB.seq.*
8: /cgn2_6/pdata/1/pubnpa/US08_PUBCOMB.seq.*
9: /cgn2_6/pdata/1/pubnpa/US09A_PUBCOMB.seq.*
10: /cgn2_6/pdata/1/pubnpa/US09B_PUBCOMB.seq.*
11: /cgn2_6/pdata/1/pubnpa/US09C_PUBCOMB.seq.*
12: /cgn2_6/pdata/1/pubnpa/US09_NEW_PUB.seq.*
13: /cgn2_6/pdata/1/pubnpa/US10A_PUBCOMB.seq.*
14: /cgn2_6/pdata/1/pubnpa/US10B_PUBCOMB.seq.*
15: /cgn2_6/pdata/1/pubnpa/US10C_NEW_PUB.seq.*
16: /cgn2_6/pdata/1/pubnpa/US10_PUBCOMB.seq.*
17: /cgn2_6/pdata/1/pubnpa/US60_NEW_PUB.seq.*
18: /cgn2_6/pdata/1/pubnpa/US60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

| Result No. | %     |       | SUMMARIES |    |                    |                   |  | Description |
|------------|-------|-------|-----------|----|--------------------|-------------------|--|-------------|
|            | Score | Match | Length    | DB | ID                 |                   |  |             |
| 1          | 258   | 100.0 | 477       | 9  | US-09-934-249-4    | Sequence 4, Appli |  |             |
| 2          | 258   | 100.0 | 1228      | 8  | US-08-825-486-9    | Sequence 9, Appli |  |             |
| 3          | 258   | 100.0 | 1228      | 8  | US-08-870-434-5    | Sequence 5, Appli |  |             |
| 4          | 258   | 100.0 | 1228      | 9  | US-09-372-044-9    | Sequence 9, Appli |  |             |
| 5          | 258   | 100.0 | 1228      | 10 | US-09-560-150-5    | Sequence 5, Appli |  |             |
| 6          | 258   | 100.0 | 1228      | 14 | US-10-067-741-5    | Sequence 5, Appli |  |             |
| 7          | 258   | 100.0 | 1297      | 14 | US-10-106-698-1370 | Sequence 1370, Ap |  |             |
| 8          | 258   | 100.0 | 1309      | 14 | US-10-171-581-277  | Sequence 277, App |  |             |
| 9          | 258   | 100.0 | 1578      | 9  | US-09-764-877-3759 | Sequence 3759, Ap |  |             |
| 10         | 258   | 100.0 | 1578      | 9  | US-09-764-877-3801 | Sequence 3801, Ap |  |             |
| 11         | 258   | 100.0 | 1578      | 15 | US-10-242-515-3799 | Sequence 3799, Ap |  |             |
| 12         | 258   | 100.0 | 1578      | 15 | US-10-242-515-3801 | Sequence 3801, Ap |  |             |
| 13         | 256.4 | 99.4  | 1223      | 9  | US-09-860-107-3037 | Sequence 3037, Ap |  |             |
| 14         | 256.4 | 99.4  | 1223      | 10 | US-09-860-706-873  | Sequence 873, App |  |             |
| 15         | 256.4 | 99.4  | 1223      | 10 | US-09-873-319-568  | Sequence 568, App |  |             |

Sequence 3901, Ap  
Sequence 3901, Ap  
Sequence 3901, Ap  
Sequence 3901, Ap  
Sequence 3947, Ap  
Sequence 3947, Ap  
Sequence 3947, Ap  
Sequence 3947, Ap  
Sequence 3946, Ap  
Sequence 15, Appl  
Sequence 1, Appl  
Sequence 2449, Ap  
Sequence 657, App  
Sequence 1067, Ap  
Sequence 18069, A  
Sequence 2, Appl  
Sequence 4959, Ap  
Sequence 1242, Ap  
Sequence 100, App  
Sequence 715, App  
Sequence 2516, A  
Sequence 36, Appl  
Sequence 25508, A  
Sequence 145, App  
Sequence 113, App  
Sequence 113, App  
Sequence 25382, A  
Sequence 1010, Ap  
Sequence 12, Appl  
Sequence 6, Appl

## ALIGNMENTS

```

RESULT 1
US-09-934-249-4
/ Sequence 4, Application US/09934249
/ Patent No. US20020115081A1
/ GENERAL INFORMATION:
/ APPLICANT: Lee, Richard T.
/ APPLICANT: Landschulz, Katherine T.
/ APPLICANT: Turi, Thomas G.
/ APPLICANT: Thompson, John F.
/ APPLICANT: Kennedy, Scott P.
/ TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
/ TITLE OF INVENTION: CARDIOVASCULAR CONDITIONS
/ FILE REFERENCE: 50738/7001/ERP/KA
/ CURRENT APPLICATION NUMBER: US/09/934,249
/ CURRENT FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: US 60/227,159
/ PRIOR FILING DATE: 2000-08-22

```

```
Query Match      100.0%; Score 258; DB 9; Length 477;
Best Local Similarity 100.0%; Pred. No. 4.4e-66;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

[illegible]



Db 271 CTCACATCGTCTTCCAGATCCCTGATGCTGAAGAGGGTGTCCGGGGCCCTGCT 330  
QY 121 CCAGAGGAGCGCCCTAACGCGCCGATCCCTGCGCGCCGCCCTGCTTCCCGGTCCTCGAG 180  
Db 331 CCAGAGGAGCGCCCTAACGCGCCGATCCCTGCGCGCCGCCCTGCTTCCCGGTCCTCGAG 390  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTCAGACATTTCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTCAGACATTTCTCCAG 450  
QY 241 CAACACCGCGCGCTTC 258  
Db 451 CAACACCGCGCGCTTC 468

RESULT 4  
US-09-372-044-9  
; Sequence 9, Application US/09372044A  
; Patent No. US20020102603A1  
; GENERAL INFORMATION:  
; APPLICANT: Dean FALB et al.  
; TITLE OF INVENTION: Compositions and Methods for the  
; TITLE OF INVENTION: Treatment and Diagnosis of Cardiovascular Disease  
; FILE REFERENCE: 7853-152  
; CURRENT APPLICATION NUMBER: US/09/372,044A  
; CURRENT FILING DATE: 1999-08-11  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 9  
; LENGTH: 1228  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(468)  
US-09-372-044-9

Query Match 100.0%; Score 258; DB 9; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 4.7e-66;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCGAGCTGCAGATCCCTGAGGAGCCGAGACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 211 GTCGGCGCGAGCTGCAGATCCCTGAGGAGCCGAGACCCAGCCAAAGGCTTCTTTCTGCTG 270  
QY 61 CTCACATCGTCTTCCAGATCCCTGATGCTGAGAGGGGTGCGCGCGCCCTGCT 120  
Db 271 CTCACATCGTCTTCCAGATCCCTGATGCTGAGAGGGGTGCGCGCGCCCTGCT 330  
QY 121 CCAGAGGAGCGCCCTAACGCGCCGATCCCTGCGCGCCGCCCTGCTTCCCTCGAG 180  
Db 331 CCAGAGGAGCGCCCTAACGCGCCGATCCCTGCGCGCCGCCCTGCTTCCCTCGAG 390  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTCAGACATTTCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTCAGACATTTCTCCAG 450  
QY 241 CAACACCGCGCGCTTC 258  
Db 451 CAACACCGCGCGCTTC 468

RESULT 5  
US-09-560-150-5  
; Sequence 5, Application US/09560150  
; Publication No. US20030073076A1  
; GENERAL INFORMATION:  
; APPLICANT: FALB, Dean A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE  
; TITLE OF INVENTION: TREATMENT AND DIAGNOSIS OF CARDIOVASCULAR DISEASE  
; FILE REFERENCE: 7853-126  
; CURRENT APPLICATION NUMBER: US/09/560,150  
; CURRENT FILING DATE: 2000-04-28

; PRIOR APPLICATION NUMBER: 09/126,540  
; PRIOR FILING DATE: 1998-07-30  
; PRIOR APPLICATION NUMBER: 08/870,434  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 08/799,910  
; PRIOR FILING DATE: 1997-02-13  
; PRIOR APPLICATION NUMBER: 60/011,787  
; PRIOR FILING DATE: 1996-02-16  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1228  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-560-150-5  
Query Match 100.0%; Score 258; DB 10; Length 1228;  
Best Local Similarity 100.0%; Pred. No. 4.7e-66;  
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GTCGGCGCGAGCTGCAGTCCGAGGAGCCGAGACCCAGCCAAAGGCTTCTTTCTGCTG 60  
Db 211 GTCGGCGCGAGCTGCAGTCCGAGGAGCCGAGACCCAGCCAAAGGCTTCTTTCTGCTG 270  
QY 61 CTCACATCGTCTTCCAGATCCCTGATGCTGAGAGGGGTGCGCGCGCCCTGCT 120  
Db 271 CTCACATCGTCTTCCAGATCCCTGATGCTGAGAGGGGTGCGCGCGCCCTGCT 330  
QY 121 CCAGAGGAGCGCCCTAACGCGCCGATCCCTGCGCGCCGCCCTGCTTCCCTCGAG 180  
Db 331 CCAGAGGAGCGCCCTAACGCGCCGATCCCTGCGCGCCGCCCTGCTTCCCTCGAG 390  
QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTCAGACATTTCTCCAG 240  
Db 391 CCCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCCTCAGACATTTCTCCAG 450  
QY 241 CAACACCGCGCGCTTC 258  
Db 451 CAACACCGCGCGCTTC 468

RESULT 6  
US-10-067-741-5  
; Sequence 5, Application US/10067741  
; Publication No. US20030097668A1  
; GENERAL INFORMATION:  
; APPLICANT: Dean A. Falb  
; APPLICANT: Katherine Galvin  
; APPLICANT: Michael Donovan  
; APPLICANT: Dennis Huszar  
; APPLICANT: Michael A. Gimbrone, Jr.  
; TITLE OF INVENTION: Compositions and Methods for the Treatment and  
; TITLE OF INVENTION: Diagnosis of  
; TITLE OF INVENTION: Cardiovascular Disease  
; FILE REFERENCE: 7853-140-999  
; CURRENT APPLICATION NUMBER: US/10/067,741  
; CURRENT FILING DATE: 2002-02-08  
; PRIOR APPLICATION NUMBER: US/09/288,292  
; PRIOR FILING DATE: 1999-04-08  
; PRIOR APPLICATION NUMBER: 08/870,434  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 08/799,910  
; PRIOR FILING DATE: 1997-02-13  
; PRIOR APPLICATION NUMBER: 60/011,787  
; PRIOR FILING DATE: 1996-02-16  
; PRIOR APPLICATION NUMBER: 08/485,573  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 08/386,844  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1228

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(468)
US-10-067-741-5

Query Match
Best Local Similarity 100.0%; Score 258; DB 14; Length 1228;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGTGCCTGCTGAGGACCGAAGCCGAGCCGAAAGGCTTCCTTTCTGCTG 60
DB 211 GTCCGGCGCCAGTGCCTGCTGAGGACCGAAGCCGAGCCGAAAGGCTTCCTTTCTGCTG 270
QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCT 120
DB 271 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCT 330
QY 121 CCAGAGGACGCCCTTAACGGCGCATCCTGGCGCCGCCACCCCTGTGTCCCGGCTTCGAG 180
DB 331 CCAGAGGACGCCCTTAACGGCGCATCCTGGCGCCGCCACCCCTGTGTCCCGGCTTCGAG 390
QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCAGCACTTTCTTCGAG 240
DB 391 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCAGCACTTTCTTCGAG 450
QY 241 CAACACCGCGCGCTTC 258
DB 451 CAACACCGCGCGCTTC 468

RESULT 7
US-10-106-698-1370
; Sequence 1370, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA00591
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 1370
; LENGTH: 1297
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1280)...(1280)
; OTHER INFORMATION: n equals a,t,g, or c
US-10-106-698-1370

Query Match
Best Local Similarity 100.0%; Score 258; DB 14; Length 1297;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGTGCCTGCTGAGGACCGAAGCCGAGCCGAAAGGCTTCCTTTCTGCTG 60
DB 242 GTCCGGCGCCAGTGCCTGCTGAGGACCGAAGCCGAGCCGAAAGGCTTCCTTTCTGCTG 301
QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCT 120
DB 302 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCT 361
QY 121 CCAGAGGACGCCCTTAACGGCGCATCCTGGCGCCGCCACCCCTGTGTCCCGGCTTCGAG 180

```

```

DB 362 CCAGAGGACGCCCTTAACGGCGCATCCTGGCGCCGCCACCCCTGTGTCCCGCTCTCCAG 421
QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCAGCACTTTCTTCGAG 240
DB 422 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCAGCACTTTCTTCGAG 481
QY 241 CAACACCGCGCGCTTC 258
DB 482 CAACACCGCGCGCTTC 499

RESULT 8
US-10-171-581-277
; Sequence 277, Application US/10171581
; Publication No. US20030104426A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: Linsley, Peter
; TITLE OF INVENTION: Signature Genes in Chronic Myelogenous Leukemia
; FILE REFERENCE: 9301-157-999
; CURRENT APPLICATION NUMBER: US/10/171,581
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,914
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 366
; SEQ ID NO 277
; LENGTH: 1309
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: AF039067
; DATABASE ENTRY DATE: 2001-06-18
US-10-171-581-277

Query Match
Best Local Similarity 100.0%; Score 258; DB 14; Length 1309;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTCCGGCGCCAGTGCCTGCTGAGGACCGAAGCCGAGCCGAAAGGCTTCCTTTCTGCTG 60
DB 322 GTCCGGCGCCAGTGCCTGCTGAGGACCGAAGCCGAGCCGAAAGGCTTCCTTTCTGCTG 381
QY 61 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCT 120
DB 382 CTCACATCGTCTTCTGCCAGATCCTGATGGCTGAAGAGGGTGTGCCGGCGCCCTGCT 441
QY 121 CCAGAGGACGCCCTTAACGGCGCATCCTGGCGCCGCCACCCCTGTGTCCCGCTTCGAG 180
DB 442 CCAGAGGACGCCCTTAACGGCGCATCCTGGCGCCGCCACCCCTGTGTCCCGCTTCGAG 501
QY 181 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCAGCACTTTCTTCGAG 240
DB 502 CCTTTAATCTGACTTCGGAGCCCTCGGACTACGCTCTGAGCTCAGCACTTTCTTCGAG 561
QY 241 CAACACCGCGCGCTTC 258
DB 562 CAACACCGCGCGCTTC 579

RESULT 9
US-09-764-877-3799
; Sequence 3799, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031

```





QY 121 CCAGAGAGCCGCTTACCGCCGATCCCTGGGCGCCACCCCTGTCTCCCGCTCTCTCGAG 180  
 Db 349 CCAGAGAGCCGCTTACCGCCGATCCCTGGGCGCCACCCCTGTCTCCCGCTCTCTCGAG 408  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTCAGACACTTCTCTCCAG 240  
 Db 409 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTCAGACACTTCTCTCCAG 468  
 QY 241 CAACACCCGCGCCCTTC 258  
 Db 469 CAACACCCGCGCCCTTC 486

RESULT 15

US-09-873-319-568  
 ; Sequence 568, Application US/09873319A  
 ; Publication No. US20030134324A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Munger, William E.  
 ; APPLICANT: Kulkarni, Prakash  
 ; APPLICANT: Getzenberg, Robert H.  
 ; APPLICANT: Waga, Iwao  
 ; APPLICANT: Yamamoto, Jun  
 ; TITLE OF INVENTION: Identifying Drugs for and Diagnosis of Benign Prostatic  
 ; TITLE OF INVENTION: Hyperplasia Using Gene Expression Profiles  
 ; FILE REFERENCE: 44921-5029-US  
 ; CURRENT APPLICATION NUMBER: US/09/873,319A  
 ; CURRENT FILING DATE: 2001-06-05  
 ; EARLIER APPLICATION NUMBER: US 60/223,323  
 ; EARLIER FILING DATE: 2000-08-07  
 ; NUMBER OF SEQ ID NOS: 755  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 568  
 ; LENGTH: 1223  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: Genbank Accession No. US20030134324A1 S81914  
 US-09-873-319-568

Query Match 99.4%; Score 256.4; DB 10; Length 1223;  
 Best Local Similarity 99.6%; Pred. No. 1.4e-65;  
 Matches 257; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GTCCGGCGCCAGTCCGAGTCGAGGACCGACCCAGCCCAAGGCTTCTTCTCTG 60  
 Db 229 GTCCGGCGCCAGTCCGAGTCGAGGACCGACCCAGCCCAAGGCTTCTTCTG 288  
 QY 61 CTCACCATGCTCTTCTGCGAGATCCTGATGGTGAAGAGGTGTGCGGCGCCCTG 120  
 Db 289 CTCACCATGCTCTTCTGCGAGATCCTGATGGTGAAGAGGTGTGCGGCGCCCTG 348  
 QY 121 CCAGAGAGCCGCTTACCGCCGATCCCTGGGCGCCACCCCTGTCTCCCGCTCTCGAG 180  
 Db 349 CCAGAGAGCCGCTTACCGCCGATCCCTGGGCGCCACCCCTGTCTCCCGCTCTCGAG 408  
 QY 181 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTCAGACACTTCTCTCCAG 240  
 Db 409 CCCTTTAATCTGACTTCGGAGCCCTCGGACTAGCTCTGGACCTCAGACACTTCTCTCCAG 468  
 QY 241 CAACACCCGCGCCCTTC 258  
 Db 469 CAACACCCGCGCCCTTC 486

Search completed: March 3, 2004, 20:45:12  
 Job time : 221 secs

US 1020979207P1



Creation date: 03-11-2004  
Indexing Officer: SMOHAMMED - SUAD MOHAMMED  
Team: 1600PrintWorkingFolder  
Dossier: 10209792

Legal Date: 03-03-2004

| No. | Doccode | Number of pages |
|-----|---------|-----------------|
| 1   | SRNT    | 2               |

Total number of pages: 2

Remarks:

Order of re-scan issued on .....